THE *MYCOBACTERIUM TUBERCULOSIS* ESX-3 SECRETION SYSTEM INTERACTOME

by
Mae Newton-Foot

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Medical Biochemistry at the Faculty of Health Sciences, University of Stellenbosch

Supervisor: Prof. Nicolaas Claudius Gey van Pittius
Department of Biomedical Science
Co-supervisor: Prof. Robin Mark Warren
Department of Biomedical Science

March 2010

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the authorship owner thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2010

Abstract

Mycobacterium tuberculosis is the causative agent of tuberculosis, a disease which causes approximately 2 million deaths each year. Despite extensive research on tuberculosis and *M. tuberculosis*, little is understood of the mechanisms of pathogenicity of the organism. The genome of *M. tuberculosis* contains five ESAT-6 gene cluster regions, each of which contains genes encoding proteins involved in the formation of a dedicated protein secretion system. Included in these regions are genes encoding exported T-cell antigens, serine proteases, ATP-binding proteins and other membrane-associated proteins. Although it is known that some of these secretion systems are involved in virulence and phagosomal escape of *M. tuberculosis*, and that deletion thereof causes attenuation of the organism, the structure, substrates and functions of the systems are largely unknown. Understanding the structure of the ESX secretion systems will advance our understanding of the mechanisms of mycobacterial pathogenicity and provide clues to ways in which to interfere with these virulence mechanisms.

The ESAT-6 gene cluster region 3, encoding the ESX-3 secretion machinery, is the only ESAT-6 gene cluster region which is essential for the *in vitro* growth of *M. tuberculosis.* It is however not required for the growth of the saprophytic mycobacterium *M. smegmatis*. In this study we have identified protein-protein interactions within the ESX-3 secretion system, using the Mycobacterial – Protein Fragment Complementation (M-PFC) mycobacterial two-hybrid system, and created a model of the *M. tuberculosis* ESX-3 secretion system. According to this model, the EsxG-EsxH and PE5-PPE4 substrate protein complexes bind to the same components of the ESX-3 secretion machinery and are secreted via the same mechanism. A knock-out of the ESX-3 secretion system in *M. smegmatis* was generated by homologous recombination to allow further research into the functions and properties of this secretion system. This knock-out was used, together with wild-type *M. smegmatis,* to investigate the secretion of the *M. tuberculosis* EsxH protein by the *M. smegmatis* ESX-3 secretion system.

The ESX-3 secretion system interactome may serve as a model for the ESX secretion systems and assist in our understanding of this secretion machinery which is key to the virulence and survival of *M. tuberculosis* and other pathogenic mycobacteria. Improved understanding of these mechanisms and their role in pathogenicity and survival may provide means of interfering with the secretion machinery, potentially leading to developments in the prevention and treatment of tuberculosis disease.

Abstrak

Tuberkulose, wat veroorsaak word deur *Mycobacterium tuberculosis,* eis jaarliks ongeveer 2 miljoen lewens. Ten spyte van uitgebreide navorsing oor tuberkulose en *M. tuberculosis* is min bekend oor die meganismes van patogenisiteit van díe organisme. Die genoom van *M. tuberculosis* bevat vyf ESAT-6 geen groep gebiede wat elk proteïene kodeer wat ‘n toegewyde sekresie sisteem vorm. Ingesluit in elk van díe geen groep gebiede is gene wat T-sel antigene, serien proteases, ATP-bindingsproteïene en ander membraan-geassosieërde proteïene kodeer. Alhoewel dit bekend is dat sekere van hierdie sekresie sisteme betrokke is by virulensie en fagosoom-ontsnapping, en dat delesie daarvan die organisme attenueer, is die struktuur, substrate en funksies van die sisteme grootliks onbekend. Kennis van die struktuur van die ESX sekresie sisteme sal ons verstaan van die meganismes van mikobakteriele patogenisiteit verbeter en leidrade verskaf na maniere om in te meng by díe meganismes van virulensie.

Die ESAT-6 geen groep gebied 3, wat die ESX-3 sekresie sisteem kodeer, is die enigste ESAT-6 geen groep gebied wat noodsaaklik is vir die *in vitro* groei van *M. tuberculosis.* Dit is egter nie nodig vir die groei van die saprofitiese mikobakterium *M. smegmatis* nie. In hierdie studie het ons proteïen-proteïen interaksies van die ESX-3 sekresie sisteem geïdentifiseer, deur middel van die Mikobakteriële - Proteïen Fragment Komplementasie (M-PFC) mikobacteriële twee-hibriede stelsel. Die interaksies is gebruik om ‘n model van die *M. tuberculosis* ESX-3 sekresie sisteem te skep. Volgens díe model bind die EsxG-EsxH en PE5-PPE4 substraat proteïen komplekse aan dieselfde komponente van die ESX-3 sekresie apparaat en word deur dieselfde meganisme uitgevoer. ‘n uitklopmutant van die ESX-3 sekresie sisteem word deur homoloë rekombinasie in *M. smegmatis* gegenereer om verdere ondersoeke na die funksies en eienskappe van hierdie sekresie sisteem in staat te stel. Hierdie uitklopmutant is tesame met die wilde-tipe *M. smegmatis* gebruik om die sekresie van die *M. tuberculosis* EsxH proteïen deur die *M. smegmatis* ESX-3 sekresie sisteem te ondersoek.

Die ESX-3 sekresie sisteem interaktoom kan dien as ‘n model vir die ESX sekresie sisteme om te help om ons kennis van hierdie sekresie apparaat, wat belangrik is vir die virulensie en oorlewing van *M. tuberculosis* en ander patogeniese mikobakterieë, te verbeter. Kennis van hierdie meganismes en hul rol in patogenisiteit en oorlewing mag maniere verskaf om by díe sekresie sisteme in te meng, wat moontlik kan lei tot ontwikkelings in die voorkoming en behandeling van tuberkulose.

Contents Page

Acknowledgements

Nico, Rob and Paul for creating such a stimulating scientific environment and for your encouragement, guidance, support and the opportunities you have afforded me to develop scientifically and personally.

Everyone in Lab 424 and the Division of Molecular Biology and Human Genetics, who have assisted me with my research and supported me throughout my studies.

My family and friends, for your unwavering support, encouragement and love.

The Harry Crossley Foundation, Ernst and Ethel Erikson Trust, Division of Molecular Biology and Human Genetics - DST/NRF Centre of Excellence in Biomedical Tuberculosis Research and Stellenbosch University for bursaries.

This research was funded by the DST/NRF Centre of Excellence in Biomedical Tuberculosis Research at Stellenbosch University, and by a grant from SATBAT – a South African/US research training collaboration – funded by the Fogarty International Center (grant: IU2RTW007370-01A1).

List of Abbreviations

aa	amino acid
ABC	ATP binding cassette
ADP	adenosine diphosphate
amp	ampicillin
Amp^R	ampicillin resistance
APS	ammonium persulphate
ATP	adenosine triphosphate
ATPase	adenosine triphosphatase
attP	attachment site of phage
BCG	Bacille Calmette et Guérin
bp	base pair
BSA	bovine serum albumin
C-	carboxy-
CAF	Central Analytical Facility
cam	Chloramphenicol
CDD	Conserved Domain Database
CF	culture filtrate
CFP-10	culture filtrate protein 10
cosA	cohesive end site A
cosB	cohesive end site B
CSU	Colorado State University
DCO	double cross-over
DHFR	dihydrofolate reductase
DNA	dioxyribonucleic acid
ecc	esx conserved component
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ESAT-6	early secretory antigenic target of 6 kDa
esp	ESX-1 secretion-associated protein
EssC	ESAT-6 secretion system C
ESX	ESAT-6 secretion system
ESX-3 KO	ESAT-6 gene cluster region 3 knock-out
$ESX\text{-}3_{MS}$	ESAT-6 gene cluster region 3 of *M. smegmatis*
Fur	ferric uptake regulator
GC	guanine and cytosine
GTP	guanine triphosphate
His	histidine
HIV	Human Immunodeficiency Virus

HRP	horseradish peroxidase
hyg	hygromycin
Hyg^R	hygromycin resistance
IDT	Integrated DNA Technologies
IdeR	iron dependant repressor
IM	inner (or plasma) membrane
int	integrase
IPTG	isopropyl-β-D-thiogalactopyranoside
kan	kanamycin
Kan^R	kanamycin resistance
kb	kilobases
KCl	potassium chloride
kDa	kiloDalton
KPL	Kirkegaard & Perry Laboratories
kV	kiloVolt
lacZ	β-galactosidase gene
LB	Luria Bertani broth
M.	*Mycobacterium*
mDHFR	murine dihydrofolate reductase
MDR	multidrug resistant
$MgCl_2$	magnesium chloride
ml	milliliter
MM	mycomembrane
M-PFC	Mycobacterial – Protein Fragment Complementation
MPTR	major polymorphic tandem repeat
MycP	mycosin protease
N-	amino-
NEB	New England Biolabs
ng	nanogram
NTP	nucleotide triphosphate
OD	optical density
oriE	*E. coli* origin of replication
oriM	mycobacterial origin of replication
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PE	proline-glutamic acid (mycobacterial protein family)
PGRS	polymorphic GC-rich sequence (mycobacterial protein family)
PPE	proline-proline-glutamic acid (mycobacterial protein family)
PPW	proline-proline-tryptophan (mycobacterial protein family)
RD	region of difference

pm	revolutions per minute
sacB	levansucrase gene
SAP	shrimp alkaline phosphatase
SCO	single cross-over
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate - polyacrylamide gel electrophoresis
sec	general secretion machinery
SOB	super optimal broth
SOC	super optimal catabolite repression
SVP	serine-valine-proline (mycobacterial protein family)
TAE	tris-acetic acid-EDTA buffer
TEMED	N,N,N',N'-tetramethylethylenediamine
tet	tetracycline
Tm	annealing temperature
trim	trimethoprim
tris	tris(hydroxymethyl)aminomethane
Tween-20	polyoxyethylene sorbitan monolaurate
Tween-80	polyoxyethylene sorbitan monooleate
T7SS	Type-VII secretion system
uF	microFarad
uFd	microFaraday
ug	microgram
ul	microliter
um	micron
UV	ultraviolet
V	volt
WCL	whole cell lysate
WT	wild-type
WXG	tryptophan-X-glycine (mycobacterial protein family)
X	variable amino acid
XDR	extensively drug resistant
X-gal	5-bromo-4-chloro-3-indolyl-β-galactoside
ZN	Ziehl-Neelsen
Zur	zinc uptake regulator
β	Beta
℃	degrees Celcius
Ω	Ohm
7H9	Middlebrook 7H9 Broth
7H11	BBL™ Seven H11 Agar Base

List of Tables

List of Figures

INTRODUCTION

The *Mycobacterium tuberculosis* ESX-3 secretion system interactome

Mycobacterium tuberculosis is the causative agent of tuberculosis, causing almost 10 million new cases of disease and resulting in approximately 1.7 million deaths each year (World Health Organisation, 2009). Despite the availability of anti-tuberculosis drugs and the BCG vaccine, the tuberculosis prevalence continues to increase (World Health Organisation, 2009). Although tuberculosis disease has been extensively researched, little is understood of the mechanisms of pathogenicity of the organism. Understanding these virulence mechanisms may lead to novel developments in the treatment and prevention of tuberculosis disease.

Comparative genomic analyses have identified a single region of the *M. tuberculosis* genome, named RD1, which is absent from all strains of the attenuated vaccine strain, *M. bovis* BCG (Mahairas *et al.*, 1996; Behr *et al.*, 1999). Deletion of RD1 has been shown to attenuate *M. tuberculosis* (Guinn *et al.*, 2004; Hsu *et al.*, 2003; Lewis *et al.*, 2003). RD1 contains nine *M. tuberculosis* genes which form part of the larger ESAT-6 gene cluster region 1. There are 5 copies of the ESAT-6 gene cluster in the *M. tuberculosis* genome, named regions 1 to 5, each encoding potent exported T-cell antigens (ESAT-6 and CFP-10), serine proteases, ATP-binding proteins and other membrane associated proteins (Tekaia *et al.*, 1999; Gey van Pittius *et al.*, 2001). Each of these ESAT-6 gene clusters encodes a dedicated protein secretion system, called ESX-1, 2, 3, 4 and 5 (Abdallah *et al.,* 2007; Simeone *et al.,* 2009).

These secretion systems are responsible for the secretion of proteins, including the T-cell antigens, ESAT-6 and CFP-10 across the mycobacterial mycomembrane (Abdallah *et al.*, 2007). Although it is known that these secretion systems are involved in virulence and phagosomal escape of *M. tuberculosis*, and that deletion of some of these regions causes attenuation of the organism, the structure, substrates and functions of the systems are largely unknown.

The ESAT-6 gene cluster region 3, encoding the ESX-3 secretion system, is the only ESAT-6 gene cluster that is essential for growth in *M. tuberculosis* (Sassetti *et al*., 2003). It is, however, not essential in the fast-growing saprophyte, *M. smegmatis*. Expression of ESX-3 is regulated by iron and zinc availability as part of the IdeR and Fur/Zur regulons and may be involved in divalent cation homeostasis (Rodriguez *et al*., 2002; Maciag *et al*., 2007).

This study investigates ESX-3 by identifying protein-protein interactions between the ESX-3 components, in order to create a model of the ESX-3 secretion system interactome. In addition, a ESX-3 knockout strain of *M. smegmatis* was constructed in this study. This will enable future studies to investigate the functions of ESX-3, which may help to establish the essential nature of this secretion system in *M. tuberculosis.* In this study the *M. smegmatis* ESX-3 knockout was used to investigate the functional conservation of ESX-3 in *M. tuberculosis* and *M. smegmatis* by establishing whether *M. smegmatis* ESX-3 is able to secrete the *M. tuberculosis* ESX-3 protein EsxH (TB10.4).

This study assists in elucidating the structure of the ESX-3 secretion systems, which serves as a model for the structure of the other ESX secretion systems, and assists in expanding our knowledge of the function of ESX-3. This study lays the foundation for future work on these systems, which may provide clues as to how these secretion mechanisms may be interfered with, potentially leading to developments in the prevention and treatment of tuberculosis disease.

PROBLEM STATEMENT

M. tuberculosis is the causative agent of tuberculosis, a disease which continues to spread and kill millions of people each year. Although the disease, and *M. tuberculosis,* has been extensively studied, little is understood of the mechanisms of pathogenicity of the organism. The five ESAT-6 gene clusters of *M. tuberculosis* each encode a dedicated secretion system, responsible for the transport of proteins from the cell. These secretion systems are essential for the virulence and survival of *M. tuberculosis*. A better understanding of the structure, substrates, mechanism, regulation and functions of these secretion systems may lead to novel developments in the treatment and prevention of tuberculosis disease.

AIMS

This study aims to investigate the *M. tuberculosis* ESX-3 secretion system, specifically

1. To identify protein-protein interactions within the ESX-3 secretion system.
2. To create a model of the ESX-3 secretion machinery.
3. To identify protein-protein interactions between the Esx proteins encoded by the ESX-3 and duplicated from it.
4. To create an ESX-3 knock-out strain of *M. smegmatis.*
5. To determine whether the ESX-3 secretion system of *M. tuberculosis* is functionally conserved in *M. smegmatis* by establishing whether the *M. smegmatis* ESX-3 is able to secrete the *M. tuberculosis* EsxH protein.

CHAPTER 1

The *Mycobacterium tuberculosis* ESX secretion systems

LITERATURE REVIEW

Tuberculosis is an infectious disease caused by the bacterium *Mycobacterium tuberculosis* (Koch, 1882). Approximately two-thirds of the world's population is infected with *M. tuberculosis*, with approximately 10 million new cases of active tuberculosis, resulting in 1.75 million deaths, each year (World Health Organisation, 2009). Despite the availability of anti-tuberculosis drugs and the BCG vaccine, the prevalence of tuberculosis disease continues to increase. The ineffectiveness of the BCG vaccine against adult tuberculosis, the high prevalence of HIV in areas with high tuberculosis burden, and the development of drug resistance, especially MDR (multidrug resistant) and XDR (extensively drug resistant) tuberculosis, contribute to the increasing tuberculosis prevalence (World Health Organisation, 2007; 2008). In order to combat this disease, new, more effective vaccines and drugs need to be developed. Despite extensive research into tuberculosis and *M. tuberculosis*, very little is understood about the mechanisms of pathogenicity of the organism. A better understanding of this pathogen and its virulence mechanisms may lead to novel developments in the treatment and prevention of tuberculosis disease.

Mycobacteria

M. tuberculosis is a member of the genus *Mycobacterium*, which consists of about 147 species and 11 subspecies, and contains both pathogenic and non-pathogenic saprophytic bacteria (Euzeby, 2009). The *Mycobacteria* are non-motile, non-sporulated, acid-fast, rod-shaped bacteria characterized by the high GC (guanine and cytosine) content of their genomes and their lipid-rich cell walls containing mycolic acids (Shinnick and Good, 1994). Mycobacteria are classified as fast-growing if they form a colony on solid growth medium within 7 days, and as slow-growing if they only form visible colonies after 7 days (Shinnick and Good, 1994). An example of a fast-growing mycobacterium is *M. smegmatis*, which forms a colony in about 3 days. *M. tuberculosis* is a slow-growing mycobacterium which only forms a colony after 3 to 4 weeks. Most fast-growing mycobacteria are saprophytic, while the majority of pathogenic mycobacteria are slow-growers.

The *M. tuberculosis* complex, comprising the closely related species *M. africanum, M. bovis, M. canetti*, *M. caprae, M. microti, M. pinnipedii, M. tuberculosis,* the oryx bacillus and the dassie bacillus, causes tuberculosis disease in humans and animals (Brosch *et al*., 2000b). The attenuated *M. bovis* BCG strain, also a member of the complex, was developed after serial passaging of a virulent *M. bovis* strain by Calmette and Guérin between 1908 and 1919 and is used to vaccinate children to prevent

tuberculosis disease (World Health Organisation, 2004). Other pathogenic mycobacteria include *M. leprae* and *M. ulcerans* which cause Leprosy and Buruli Ulcer respectively (MacCallum *et al*., 1948; Gelber, 1994). Some mycobacteria such as *M. smegmatis* occur in the environment as saprophytes, while others infect a variety of non-human hosts.

The ESAT-6 gene cluster

The whole genome sequence of the laboratory strain *M. tuberculosis* H37Rv was described by Cole *et al* (1998). This, together with the use of other comparative genomic techniques, allowed for the identification of various Regions of Difference (RDs) between closely related virulent and non-virulent strains and species (Mahairas *et al.,* 1996; Philipp *et al.,* 1996; Brosch *et al.,* 1998, 1999, 2000a; Gordon *et al.,* 1999a; Behr *et al.,* 1999). These analyses identified a single region of the *M. tuberculosis* and *M. bovis* genomes, named RD1, which is absent from all substrains of the vaccine strain *M. bovis* BCG (Mahairas *et al*., 1996; Behr *et al*., 1999). This is believed to be the principal deletion which resulted in the attenuation of *M. bovis* BCG (Behr *et al*., 1999; Brosch *et al*., 2000a). This region is present in the genomes of *M. bovis* and *M. africanum,* both virulent members of the *M. tuberculosis* complex, but is absent in *M. microti,* which seldom causes disease in immuno-competent individuals (Van Soolingen *et al.,* 1998, Gordon *et al*., 1999a; Brodin *et al*., 2002). Deletion of RD1 has been shown to cause attenuation of *M. tuberculosis* (Hsu *et al*., 2003; Lewis *et al*., 2003; Guinn *et al*., 2004), leading to the hypothesis that the RD1 region contains elements which contribute to mycobacterial pathogenicity.

RD1 is a 9505bp region containing nine *M. tuberculosis* genes, Rv3871 to Rv3879c. Included in this region are genes encoding the 6 kDa early secreted antigenic target (ESAT-6, Rv3875) and culture filtrate protein 10 (CFP-10, Rv3874) (Berthet *et al*., 1998; Gey van Pittius *et al.,* 2001). ESAT-6 and CFP-10 are two potent T-cell antigens which were identified from the short term culture filtrates of *M. tuberculosis* (Andersen *et al*., 1995). Although these proteins contain no known signal sequences (Sorensen *et al*., 1995; van Pinxteren *et al*., 2000), their secretion is essential for *M. tuberculosis* virulence (Hsu *et al*., 2003; Lewis *et al*., 2003; Stanley *et al*., 2003; Guinn *et al*., 2004).

Further analysis identified a cluster of genes surrounding and including the ESAT-6 and CFP-10 genes, of which there are 5 copies in the *M. tuberculosis* genome (Tekaia *et al*., 1999). These 5 gene clusters have been named the ESAT-6 gene cluster regions 1 (Rv3866-Rv3883c, encompassing

RD1), 2 (Rv3884c-Rv3895c), 3 (Rv0282-Rv0292), 4 (Rv3444c-Rv3450c) and 5 (Rv1782-1798) and phylogenetic analyses indicate that they were duplicated from the ancestral region, Region 4, in the order 3, 1, 2 and then 5 (Figure 1.1, Gey van Pittius NC, personal communication). Twelve gene families are represented within the 5 gene cluster regions and were designated families A to L according to their position in Region 1 (Gey van Pittius *et al*., 2001). Six of these gene families are present in all 5 regions, and encode the ESAT-6 and CFP-10 homologs as well as a transmembrane ATPase, a transmembrane ATP-binding protein, a subtilisin-like membrane-anchored cell wall-associated serine protease (mycosin) and a putative integral membrane pore protein (Tekaia *et al*., 1999; Gey van Pittius *et al*., 2001). In addition to the conserved components of the ESAT-6 gene clusters, regions 3, 1, 5 and 2 also contain genes encoding several other proteins including PE and PPE proteins. Other region-specific genes also occur in some of these clusters (Gey van Pittius *et al*., 2001). The presence and positions of the Family A to L genes in each of the 5 ESAT-6 gene cluster regions of *M. tuberculosis* are given in Table 1.1 and the proteins they encode are described below.

A systematic genetic nomenclature has been proposed for the components of these gene cluster regions, and Type VII secretion systems (T7SSs) in general (Bitter *et al.,* 2009). The newly proposed terminology is given alongside the old terms in Table 1.1 and in Figure 1.1, however previous terminology will be used throughout this thesis.

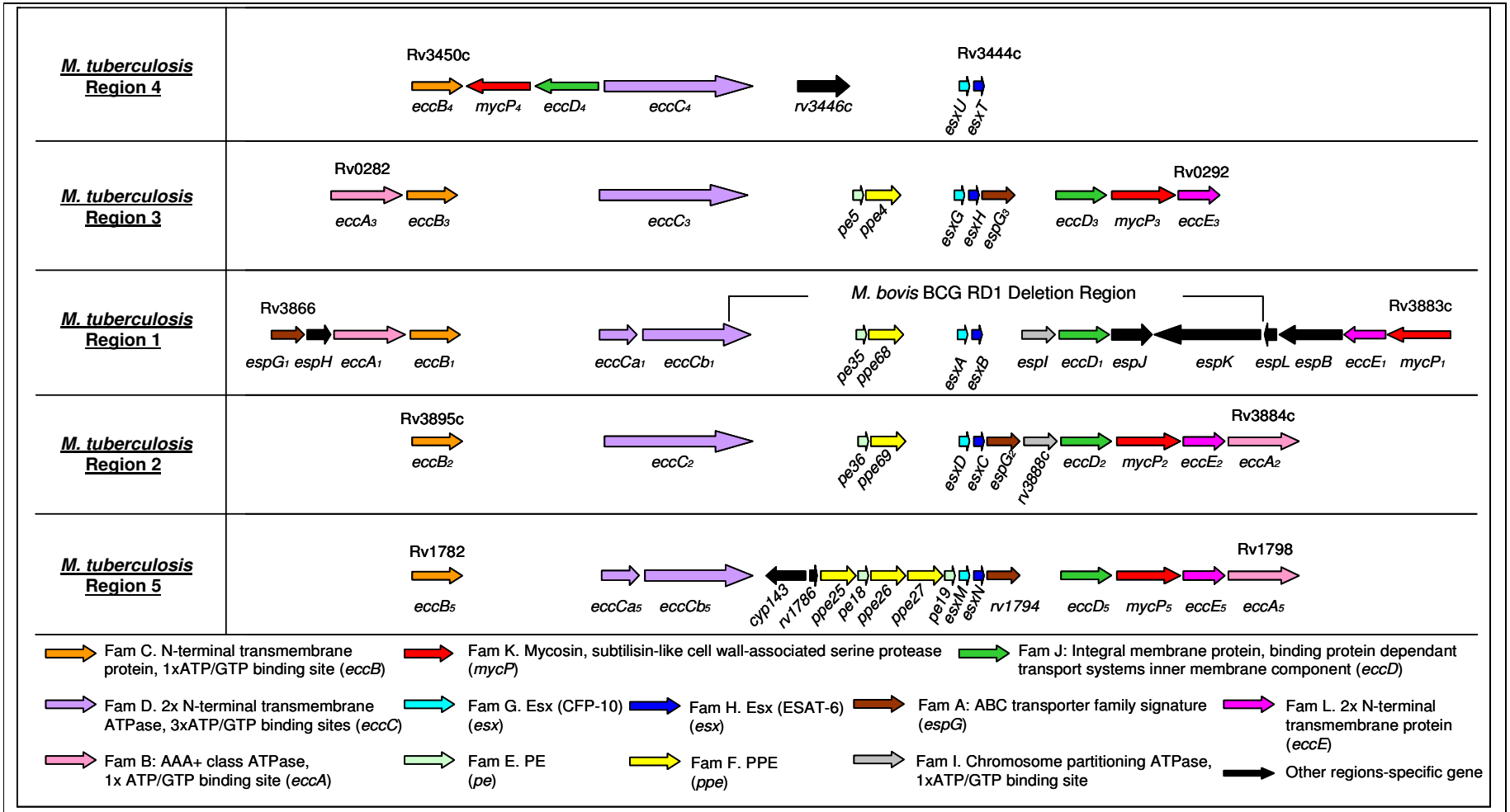


Figure 1.1. The ESAT-6 gene clusters of *M. tuberculosis*. The *M. tuberculosis* ESAT-6 gene clusters evolved in the order 4, 3, 1, 2 and then 5, through duplication events and the incorporation of additional genes. The genes of the ESAT-6 gene cluster region 4 are maintained through all the duplications and the PE/PPE genes, incorporated into Region 3, in the subsequent duplications. Adapted from Gey van Pittius *et al.* (2001)

Table 1.1. The components of the five ESAT-6 gene clusters of *M. tuberculosis.*

Gene family	Description	Presence of genes in the ESAT-6 gene cluster regions				
		1	2	3	4	5
A	ABC transporter family signature	Rv3866 (*espG$_1$*)	Rv3889c (*espG$_2$*)	Rv0289 (*espG$_3$*)		Rv1794
B	AAA+ class ATPases, CBXX/CFQX family, SpoVK, 1x ATP/GTP-binding site	Rv3868 (*eccA$_1$*)	Rv3884c (*eccA$_2$*)	Rv0282 (*eccA$_3$*)		Rv1798 (*eccA$_5$*)
C	Amino terminal transmembrane protein, possible ATP/GTP-binding motif	Rv3869 (*eccB$_1$*)	Rv3895c (*eccB$_2$*)	Rv0283 (*eccB$_3$*)	Rv3450c (*eccB$_4$*)	Rv1782 (*eccB$_5$*)
D	DNA segregation ATPase, ftsK chromosome partitioning protein, SpoIIIE, YukA, 3x ATP/GTP-binding sites, 2x amino-terminal transmembrane protein	Rv3870 (*eccCa$_1$*) -Rv3871 (*eccCb$_1$*)	Rv3894c (*eccC$_2$*)	Rv0284 (*eccC$_3$*)	Rv3447c (*eccC$_4$*)	Rv1783 (*eccCa$_5$*) -Rv1784 (*eccCb$_5$*)
E	PE	Rv3872 (*pe35*)	Rv3893c (*pe36*)	Rv0285 (*pe5*)		Rv1788 (*pe18*) Rv1791 (*pe19*)
F	PPE	Rv3873 (*ppe68*)	Rv3892c (*ppe69*)	Rv0286 (*ppe4*)		Rv1787 (*ppe25*) Rv1789 (*ppe26*) Rv1790 (*ppe27*)
G	CFP-10, Esx family protein	Rv3874 (*esxB*)	Rv3891c (*esxD*)	Rv0287 (*esxG*)	Rv3445c (*esxU*)	Rv1792 (*esxM*)
H	ESAT-6, Esx family protein	Rv3875 (*esxA*)	Rv3890c (*esxC*)	Rv0288 (*esxH*)	Rv3444c (*esxT*)	Rv1793 (*esxN*)
I	ATPases involved in chromosome partitioning, 1x ATP/GTP-binding motif	Rv3876 (*espI*)	Rv3888c			
J	Integral inner membrane protein, binding-protein-dependent transport systems inner membrane component signature, putative transporter protein	Rv3877 (*eccD$_1$*)	Rv3887c (*eccD$_2$*)	Rv0290 (*eccD$_3$*)	Rv3448 (*eccD$_4$*)	Rv1795 (*eccD$_5$*)
K	Mycosin, subtilisin-like cell wall-associated serine protease	Rv3883c (*mycP$_1$*)	Rv3886c (*mycP$_2$*)	Rv0291 (*mycP$_3$*)	Rv3449 (*mycP$_4$*)	Rv1796 (*mycP$_5$*)
L	2x amino-terminal transmembrane protein	Rv3882c (*eccE$_1$*)	Rv3885c (*eccE$_2$*)	Rv0292 (*eccE$_3$*)		Rv1797 (*eccE$_5$*)

The systematic genetic nomenclature for T7SSs (Bitter *et al.,* 2009) is given in brackets. *ecc*: esx conserved component; *esp*: ESX-1 secretion-associated protein; *mycP:* mycosin protease.

Esx family proteins – Family G (CFP-10) and Family H (ESAT-6)

ESAT-6 and CFP-10 were first identified by studies which aimed to identify immunogens which could be used to develop new, more effective anti-tuberculosis vaccines. *M. tuberculosis* culture filtrate proteins (CFPs) were purified and tested to determine their antigenicity (Andersen *et al.*, 1995). ESAT-6 and CFP-10 were highlighted by these studies due to their potent antigenicity (Sorensen *et al.*, 1995; Berthet *et al.*, 1998).

Investigation of the *M. tuberculosis* whole genome sequence identified 23 genes, including those encoding ESAT-6 and CFP-10, which encode a family of related proteins, the Esx proteins (Cole *et al.*, 1998; Tekaia *et al.*, 1999). These genes were named *esxA*-*esxW*, with their respective proteins EsxA-EsxW. Related proteins have also been identified in actinobacteria and other low GC Gram-positive bacteria (Gey van Pittius *et al.,* 2001; Pallen, 2002). They are small proteins of approximately 100 amino acids, which although not highly conserved in sequence, each contain a WXG amino acid motif (Cole *et al.*, 1998; Pallen, 2002). They form a characteristic helix-turn-helix structure (Figure 1.2), the hairpin bend formed by the WXG motif (Pallen, 2002; Renshaw *et al.,* 2005). ESAT-6 and CFP-10 are encoded by *esxA* and *esxB*, and also named EsxA and EsxB, respectively. *esxA* and *esxB* are located directly adjacent to one another in the *M. tuberculosis* genome and are cotranscribed (Berthet *et al.*, 1998). Twenty-two of the *M. tuberculosis esx* genes occur in pairs, of which five pairs occur within the ESAT-6 gene clusters. The additional six *esx* gene-pairs, and the individual *esx* gene, were shown by Gey van Pittius *et al* (2006) to be duplicated from the various ESAT-6 gene clusters. The *esx* gene pairs form operons resulting in the coexpression of the two Esx proteins, which interact to form heterodimers (Figure 1.2) which despite the absence of known secretion signals are secreted from the cell (Sorensen *et al.*, 1995; van Pinxteren *et al.*, 2000; Pym *et al.*, 2003). Although the Esx proteins have been implicated in various functions including virulence, phagosome escape, cytolysis and cytotoxicity (Hsu *et al.*, 2003; Gao *et al.*, 2004); their precise functions have not been elucidated.

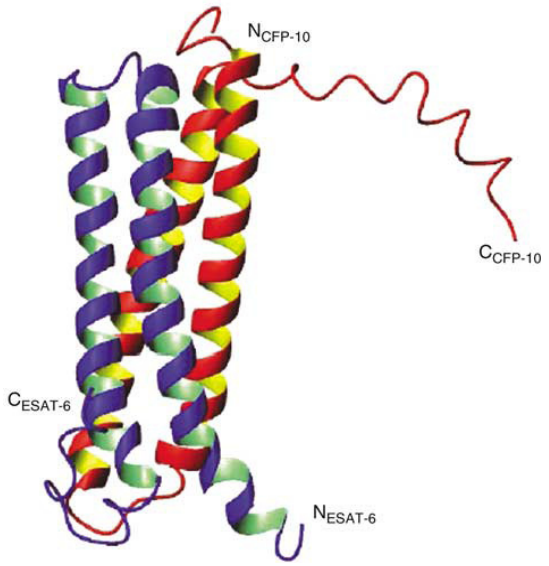


Figure 1.2. The Esx protein complex. ESAT-6 and CFP-10 interact to form a dimer, with each protein consisting of a helix-turn-helix motif and an unstructured C-terminal region. Source: Renshaw *et al* (2005).

The PE and PPE proteins – Family E (PE) and Family F (PPE)

Directly upstream of the *esx-esx* operons in the ESAT-6 gene cluster regions 1, 2, 3 and 5 is another pair of conserved genes, from the PE and PPE gene families. The presence of these genes is also conserved in four of the six duplications of the esx-esx operons outside of the ESAT-6 gene clusters, suggesting that the 4 genes were duplicated together from the ESAT-6 gene clusters (Gey van Pittius *et al*., 2006). Many additional copies of the PE and PPE genes occur in *M. tuberculosis*, with the two gene families together comprising approximately 10% of the coding material in the *M. tuberculosis* genome (Cole *et al*., 1998). *M. tuberculosis* contains 99 PE-encoding genes, which are characterized by the proline-glutamic acid (PE) motif at amino acid positions 8 and 9 in a conserved 110 amino acid N-terminal domain (Cole *et al*., 1998; Gordon *et al*., 1999b; Camus *et al*., 2002). The PPE proteins, 69 of which are encoded by *M. tuberculosis*, have a proline-proline-glutamic acid (PPE) motif at positions 7 to 9 in a unique conserved N-terminal domain of approximately 180 amino acids (Cole *et al*., 1998; Camus *et al*., 2002). The N-terminal domains vary significantly between these two protein families, which both possess highly variable C-terminal domains (Gordon *et al*., 1999b). The PE and PPE families have been further subdivided into subfamilies based on their C-terminal domains (Figure 1.3). The PE family has been divided into two subfamilies. The polymorphic GC-rich-repetitive sequence (PGRS) subfamily consists of 65 proteins with multiple tandem repeats of either glycine-glycine-

alanine or glycine-glycine-asparagine motifs in the C-terminal domain (Poulet and Cole, 1995; Gordon *et al.*, 1999b). The other PE subfamily combines 34 PE proteins with low C-terminal homology (Gordon *et al.*, 1999b). The PPEs are subdivided into 4 subfamilies (Gordon *et al.*, 1999b; Adindla and Guruprasad, 2003). The PPE-SVP subfamily comprises 24 PPE proteins with a Gly-X-X-Ser-Val-Pro-X-X-Trp motif between amino acids 300 and 350 (Adindla and Guruprasad, 2003). The PPE-MPTR subfamily of 23 members contains multiple tandem repeats of a Asp-X-Gly-X-Gly-Asn-X-Gly motif (Hermans *et al.*, 1992; Cole *et al.*, 1998). The third subfamily, the PPE-PPW subfamily, contains a conserved 44 amino acid region consisting of Phe-X-Gly-Thr and Pro-X-X-Pro-X-X-Trp motifs (Adindla and Guruprasad, 2003), while the fourth subfamily contains PPE proteins of low C-terminal homology (Gordon *et al.*, 1999b).

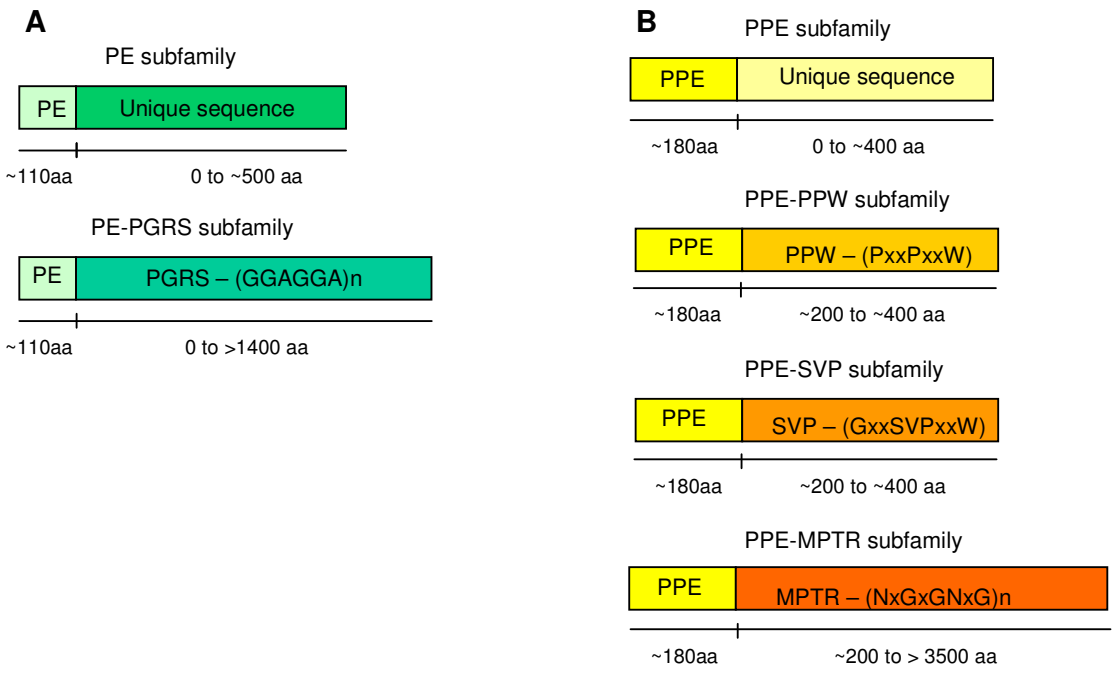


Figure 1.3. PE and PPE protein structure. Diagrammatic representation of the PE (A) and PPE (B) protein subfamilies of *M. tuberculosis,* showing the conserved N-terminal domains and the variable C-terminal sequences classifying the proteins into the various subfamilies. Adapted from Gey van Pittius *et al* (2006).

The large number of PE and PPE proteins encoded by *M. tuberculosis* suggests that these proteins must play an important role in the organism. However the functions of these protein families have yet to be elucidated. Studies have suggested that some PPE proteins may be cell wall-associated and that some are partially exposed on the cell surface of the organism (Doran *et al.*, 1992; Sampson *et*

al., 2001; Okkels *et al*., 2003; Le *et al*., 2005). Various PE-PGRS proteins have been shown to be cell surface constituents which influence colony morphology, cellular architecture and are involved in cell-cell interactions (Brennan *et al*., 2001; Banu *et al*., 2002; Delogu *et al*., 2004). Outer membrane anchoring domains have been identified in 40 PE and PPE proteins which may have the potential to form β-barrel outer membrane protein structures (Pajon *et al*., 2006) and PE35 (from the ESAT-6 gene cluster region 1) is secreted by *M. tuberculosis* (Fortune *et al*., 2005). The high degree of C-terminal variation in these protein families suggests that they may play a role in antigenic variation or in the inhibition of antigen processing (Cole *et al*., 1998; Cole, 1999; Gordon *et al*., 1999b). Alternative functions have been suggested for some PE and PPE proteins, including PPE37, expression of which is upregulated under iron-poor conditions and which has been proposed to be a siderophore-type protein (Rodriguez *et al*., 1999; Rodriguez *et al*., 2002). The PE-PGRS Wag22 has been annotated as a fibronectin-binding protein (Abou-Zeid *et al*., 1991; Espitia *et al*., 1999). Various PE and PPE proteins have also been implicated in phagosome-lysosome fusion, macrophage vacuole acidification, granuloma persistence, replication in macrophages and virulence and some have been shown to be essential for *in vitro* or *in vivo* growth (Sassetti *et al*., 2003; Li *et al*., 2005).

The presence of these proteins in the cell membrane, cell wall and culture filtrates of mycobacteria, and the various functions in which they have been implicated requires targeting of the proteins to the membrane and/or their secretion from the cell. However, akin to the Esx proteins, no known secretion signals have been identified in these proteins. PPE41 and PE25, encoded by Rv2430c and Rv2431c, are cotranscribed and interact to form a 1:1 complex (Tundup *et al*., 2006; Strong *et al.,* 2006) and several other PE-PPE protein pairs have also been predicted to form complexes (Strong *et al.,* 2006; Riley *et al.,* 2008). PPE68 (encoded from the ESAT-6 gene cluster region 1) has in addition been shown to interact with the Esx proteins EsxA, B and H (Okkels and Andersen, 2004). The physical association of the ancestral PE and PPE genes with the Esx genes, the analogy of their expression, the interactions between them and their proposed involvement in virulence and pathogenesis, suggest that the functions of the Esx and PE and PPE proteins may be linked.

The mycosins – Family K

Each of the ESAT-6 gene clusters encodes a subtilisin-like cell wall-associated serine protease (Cole *et al*., 1998; Gey van Pittius *et al*., 2001). These proteases are named mycosin-1, -2, -3, -4 and -5

according to the numbering adopted for the ESAT-6 gene clusters, encoded by *mycP1* to *mycP5* (Brown *et al.*, 2000). The mycosins contain several features of serine proteases of other bacteria. The presence of the catalytic triad, consisting of an aparagine, histidine and serine residue, together with specific active site signatures classified these proteases as subtilisin-like proteases. Other features of these proteases include their hydrophobic N-termini which are likely signal peptides, cleaved at a conserved sequence position following an Ala-X-Ala motif. The C-terminal domains consist of hydrophobic stretches interspersed with charged residues indicative of transmembrane domains and a proline-rich linker connects the transmembrane and catalytic domains. The structure of the mycosins is described in Figure 1.4. The mycosins are located in the cell wall and cell membrane of mycobacteria and may be involved in processing of extracellular or secreted proteins, and in this way contribute to virulence of mycobacteria (Brown *et al.*, 2000). Mycosin-1 has been shown to be expressed following the infection of macrophages, supporting the role of these proteases in mycobacterial pathogenicity (Dave *et al.*, 2002). The substrates and specific functions of the mycosins have not yet been determined.

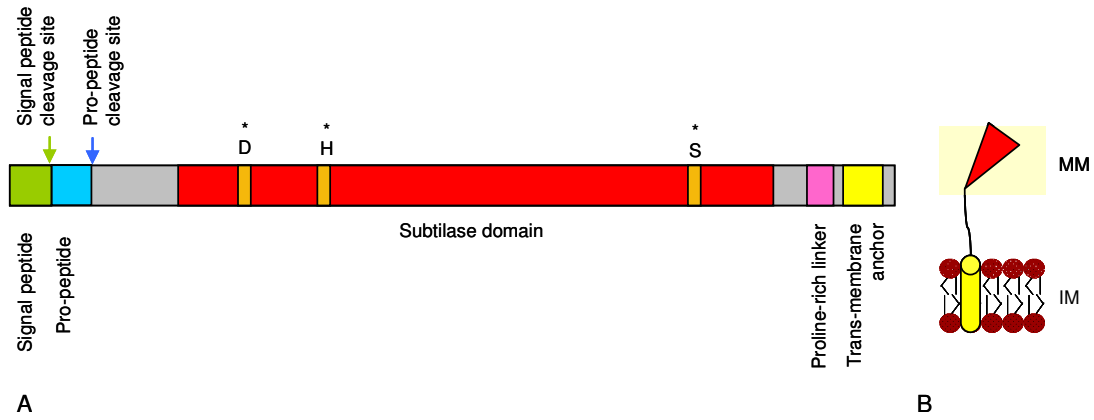


Figure 1.4. The structure of the mycosins. The predicted domains (A) and proposed structure (B) of the mycosin proteases. The mycosin proteases each contain an N-terminal signal peptide, a pro-peptide, a subtilase domain containing the catalytic triad (D-H-S)*, a proline-rich linker and a C-terminal transmembrane domain. The signal sequence targets the mycosin for secretion and is cleaved at an A-I cleavage site. The pro-peptide is cleaved at a putative Q-R pro-peptide cleavage site, resulting in activation of the subtilisin protease domain. The transmembrane domain anchors the mycosin in the cell membrane. IM: inner membrane; MM: mycomembrane.

ATPases – Families B, D and I

The ESAT-6 gene clusters each contain at least one, and up to three ATPase genes. ATPases are enzymes which convert chemical energy, in the form of ATP, to biological activity through the dephosphorylation of ATP to ADP. The ATPases are encoded by gene families B, D and I.

The Family B ATPase has been described as an AAA+ class ATPase of the CBXX/CFQX family containing one ATP/GTP-binding site (Tekaia *et al.*, 1999). AAA+ ATPases have been shown to be essential for the assembly of the bacterial Type VI secretion system machinery (Bonemann *et. al.,* 2009). A Family B ATPase is encoded by each of the ESAT-6 gene clusters duplicated from region 4, but not by region 4 (Gey van Pittius *et al.*, 2001). Rv3868, encoded by region 1, is approximately 63 kDa in size, and assembles as a hexamer (Ogura *et al.*, 2004; Luthra *et al.*, 2008). Each peptide consists of 2 domains, the C-terminal ATPase and oligomerisation domain and the helical N-terminal domain which is involved in the regulation of C-terminal ATPase activity (Figure 1.5). ATP binding to the protein and subsequent hydrolysis results in “open-close” movements of the protein domains, predicted to allow interactions with, and energy transfer to other proteins (Luthra *et al.*, 2008).

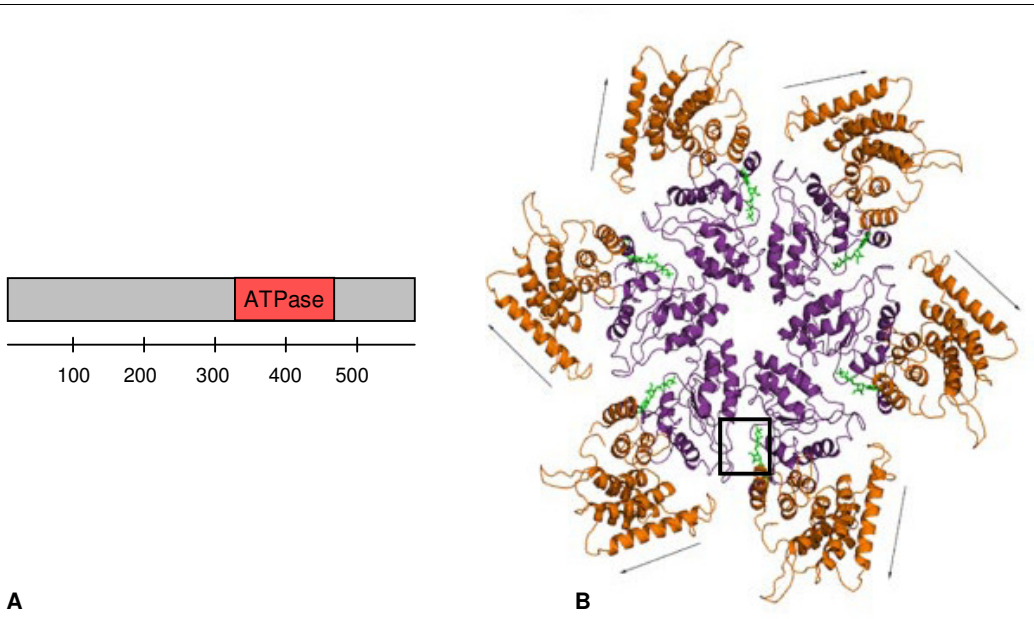


Figure 1.5. The structure of the Family B ATPase. A. The family B ATPases contain an ATPase domain of the CBXX/CFQX family in the C-terminal of the protein. B. The ATPase hexamerises. The proteins interact via the C-terminal region (in purple), which also contains the ATPase domain (blocked). ATP hydrolyses results in movements of the N-terminal region (in brown), as represented by the arrows. Source: Luthra *et al.* (2008).

The Family D ATPase is described as a DNA segregation ATPase and ftsK chromosome partitioning protein of the FtsK/SpoIIIE family and contains three ATP/GTP-binding motifs and two N-terminal transmembrane domains (Figure 1.6). Proteins of this family are essential for the functioning of Type IV secretion systems where they function as coupling proteins (Christie *et al.,* 2005). In Regions 1 and 5, the gene encoding this ATPase has been split in two resulting in the expression of a transmembrane protein containing a single ATPase domain and a cytoplasmic protein with two ATPase domains (Tekaia *et al*., 1999; Gey van Pittius *et al*., 2001).

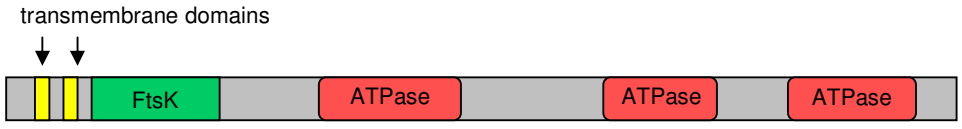


Figure 1.6. The structure of the Family D ATPase. The Family D ATPases contain 3 P-loop ATPase domains and a FtsK motif and are anchored in the membrane via 2 N-terminal transmembrane domains.

The Family I ATPase, which shows homology with ATPases involved in chromosome partitioning, is only encoded by the ESAT-6 gene cluster regions 1 and 2. This ATPase is proline and alanine rich, contains an ATP/GTP binding motif and may be membrane bound (Gey van Pittius *et al*., 2001).

Integral membrane protein – Family J

The Family J proteins, encoded by all five ESAT-6 gene cluster regions, are predicted to consist of eleven or twelve transmembrane helices which form a pore through the lipid bilayer. This protein is a putative transporter protein as it contains the signature of the inner membrane-component of binding protein-dependent transport systems (Tekaia *et al*., 1999; Gey van Pittius *et al*., 2001).

ABC transporter – Family A

The Family A genes present in the ESAT-6 gene cluster regions 1, 3, 2 and 5 encode proteins of approximately 500 amino acids which contain a hydrophobic region in the N-terminal domain (Tekaia *et al*., 1999; Gey van Pittius *et al*., 2001). An ABC (ATP binding cassette) signature has been identified in this protein and homology has been found with DNA binding proteins (Gey van Pittius *et al*., 2001; Tuberculist).

Amino-terminal membrane proteins – Families C and L

Two additional membrane proteins are encoded by the ESAT-6 gene clusters. The Family C amino-terminal transmembrane protein is encoded by all five ESAT-6 gene clusters and a possible ATP/GTP binding motif has been identified. The Family L proteins contain two amino-terminal transmembrane motifs and is encoded in regions 1, 3, 2 and 5 (Gey van Pittius *et al.*, 2001). Little else is known about these protein families.

The ESX secretion systems

Mycobacteria have a complex cell wall structure due to the presence of mycolic acids (large, hydroxylated branched-chain fatty acids) which are covalently linked to the cell wall to form an additional hydrophobic layer called the mycomembrane (Bayan *et al.*, 2003). The mycomembrane not only forms a barrier to the influx of hydrophilic substances, but also restricts the secretion of hydrophilic molecules, including extracellular proteins, from the cell. Therefore the secretion of extracellular proteins, including Esx, PE and PPE, likely requires an active secretion system. These proteins do not contain any known secretion signals and therefore do not appear to be secreted through the general secretion machinery (Sec) or other previously identified secretion mechanisms (Sorensen *et al.*, 1995; van Pinxteren *et al.*, 2000). It is proposed that the proteins encoded by each ESAT-6 secretion system form a dedicated secretion system responsible for the secretion of the Esx, PE and PPE proteins and other substrates across the mycomembrane. These secretion systems were named the ESAT-6 secretion systems 1 to 5 (ESX-1 to 5) and have been classified as a novel type of secretion system, Type-VII secretion system (Abdallah *et al.*, 2007). Several ESAT-6 gene cluster region 1 proteins have been shown to be essential for the secretion of ESAT-6 and CFP-10 affirming the hypothesis that these gene clusters encode secretion machinery (Brodin *et al.*, 2006). Each ESX secretion system is predicted to be responsible for the secretion of the Esx proteins encoded by its ESAT-6 gene cluster, and may in addition secrete the associated PE and PPE proteins as well as other unassociated proteins. Figure 1.7 describes the basic secretion mechanism of the ESX secretion systems. Each of the *M. tuberculosis* ESX secretion systems is described below.

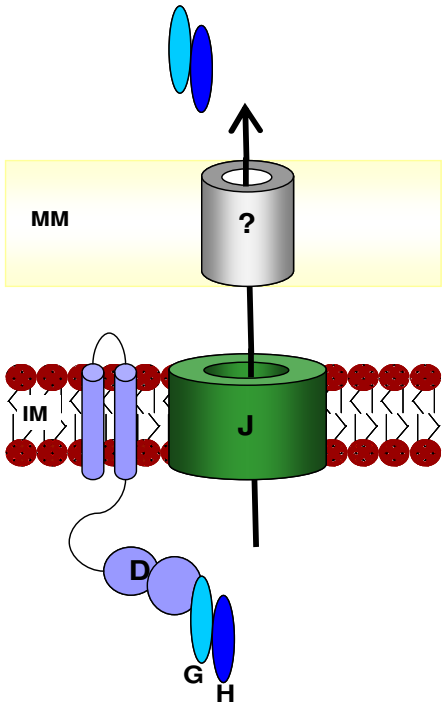


Figure 1.7. The basic Type-VII ESX secretion machinery. It is proposed that the Esx proteins (Family G and H) from each ESAT-6 gene cluster interact to form a complex which interacts with the FtsK/SpoIIE (Family D) transmembrane ATPase to provide the energy for translocation of the protein complex through a membrane pore protein, likely the Family J integral membrane protein. The functions of the other components of the ESAT-6 gene clusters remain unknown. In addition the mycomembrane channel component remains unidentified.

ESX-1

Due to its direct involvement in the virulence of *M. tuberculosis*, ESX-1 has been extensively researched and is the best characterized ESX secretion system. The involvement of the ESAT-6 gene cluster region 1 in secretion was recognized when it was noted that several genes surrounding *esxA* and *esxB* are essential for ESAT-6 and CFP-10 secretion (Hsu *et al*., 2003; Pym *et al*., 2003; Stanley *et al*., 2003; Gao *et al*., 2004; Guinn *et al*., 2004). Thus it was deduced that the RD1 region, and as such the ESAT-6 gene cluster region 1, is involved in ESAT-6 secretion.

In addition to ESAT-6 and CFP-10, the *M. tuberculosis* ESX-1 secretion system also secretes PE35, PPE68 and EspB (Rv3881c, ESX-1 secretion-associated protein B), EspF (Rv3865, ESX-1 secretion-associated protein F, encoded directly upstream of the ESAT-6 gene cluster region 1) and three genetically unlinked proteins, EspA (Rv3616c, ESX-1 secretion-associated protein A), EspC

(Rv3615c, ESX-1 secretion-associated protein C) and EspR (Rv3849, ESX-1 secretion-associated protein R) (Fortune *et al*., 2005; McLaughlin *et al*., 2007; Xu *et al*., 2007; Raghavan *et al*., 2008; Giuseppe Champion *et al*., 2009). ESAT-6 and CFP-10 are coexpressed in an operon and interact to form a heterodimer which is secreted from the cell (Berthet *et al*., 1998). The genes encoding EspA and EspC are also part of an operon with Rv3614c. Interestingly EspA, EspC and Rv3614c show significant homology to Rv3864, EspF and Rv3867, which are located directly upstream of the ESAT-6 gene cluster region 1 (Fortune *et al*., 2005) and EspC and EspF have been shown to interact (Giuseppe Champion *et al*., 2009). EspR is a transcriptional regulator which activates transcription from the Rv3616c-Rv3614c promoter ensuring proper functioning of the ESX-1 secretion machinery. EspR is also secreted by ESX-1, suggesting that ESX-1 activity is regulated by a direct negative feedback system (Raghavan *et al*., 2008). EspB is an additional ESX-1 substrate, which is encoded by the ESAT-6 gene cluster region 1 (McLaughlin *et al*., 2007). All these ESX-1 substrate proteins appear to be dependant on each other for secretion, despite variations in the specific ESX-1 components required for their individual secretion (Fortune *et al*., 2005; McLaughlin *et al*., 2007; Xu *et al*., 2007; Raghavan *et al*., 2008). It is suggested that these proteins may interact prior to, or during the secretion process and that these interactions are essential for proper ESX-1 functioning. Alternatively, these proteins may not be substrates but rather components of the ESX-1 secretion machinery, which are incidentally secreted (Fortune *et al*., 2005; Ize and Palmer, 2006). The secretion of all these ESX-1 substrates is required for full virulence of *M. tuberculosis.*

Brodin *et al* (2006) classified the gene components of the ESAT-6 gene cluster region 1 into four groups according to their roles in ESAT-6 secretion. (1) Genes required for the presence of ESAT-6 and CFP-10 in the whole cell lysate; *esxA, esxB* and *Rv3872* (encoding PE35). (2) Genes required for the secretion of ESAT-6 and CFP-10 from the cell; Rv3877, Rv3871, Rv3870, Rv3868 and Rv3869. (3) Genes which do not affect ESAT-6 and CFP-10 secretion or ESAT-6 specific immunogenicity but inactivation of which leads to enhanced virulence; *Rv3864, Rv3867, Rv3873* (encoding PPE68), *Rv3876, Rv3878* and *Rv3879*. (4) Genes which do not affect ESAT-6/CFP-10 secretion, but deletion of which causes attenuation; *Rv3865*, *Rv3866*.

The roles of some of the proteins required for ESAT-6/CFP-10 secretion have been previously described; Rv3868 and Rv3871 are ATPases which presumably provide the energy for transport of the substrates across the cell membrane. The ESAT-6/CFP-10 complex binds to Rv3871 (Renshaw *et*

al., 2002; Hsu *et al*., 2003; Stanley *et al*., 2003; Guinn *et al*., 2004; Renshaw *et al*., 2005; Champion *et al*., 2006), and EspC has been shown to bind to Rv3868 (Giuseppe Champion *et al*., 2009). Therefore it appears that these proteins are responsible for providing the energy for secretion of specific substrates. Rv3868 may also act as a chaperone, assisting in the formation of the secreted protein complex. Rv3870, which is membrane associated, binds to Rv3871, anchoring it to the membrane and may facilitate the functioning of Rv3871 (Stanley *et al*., 2003). Rv3869 is another membrane protein of unknown function, which likely forms part of the secretion machinery. Rv3877 has 11 transmembrane helices and is believed to form the membrane pore through which the substrates are transported (Tekaia *et al*., 1999; Gey van Pittius *et al*., 2001). It is suggested that PPE68 (encoded by Rv3873), may be a gating protein, controlling the secretion of ESAT-6, thereby explaining the increase in ESAT-6 secretion in Rv3873 deletion mutants (Brodin *et al*., 2006). Protein interactions between PPE68 and ESAT-6, CFP-10, Rv3868, Rv3866 and itself have been identified (Okkels and Andersen, 2004; Teutschbein *et al*., 2009). PPE68 may interact with the secretion machinery and the ESAT-6/CFP-10 complex to prevent ESX-1 secretion. Interestingly PE35, which appears to be required for the expression of ESAT-6 and CFP-10 is also secreted by *M. tuberculosis* (Fortune *et al*., 2005). The *M. tuberculosis* ESX-1 secretion machinery is described in Figure 1.8.

ESX-1 in *M. tuberculosis* appears to be involved in cytolysis, haemolysis, cytotoxicity to macrophages, bacterial spreading and macrophage escape, leading to its role in virulence (Hsu *et al*., 2003; Gao *et al*., 2004; van der Wel *et al.,* 2007). Interestingly, the non-pathogenic *M. smegmatis* contains a functionally equivalent ESX-1, which is able to secrete *M. tuberculosis* ESAT-6 and CFP-10 (Converse and Cox, 2005). *M. smegmatis* ESX-1 is involved in conjugal DNA transfer, and the *M. tuberculosis* ESX-1 is able to complement the conjugation phenotype of *M. smegmatis* ESX-1 mutants (Flint *et al*., 2004; Coros *et al*., 2008). This suggests that the function of ESX-1 is conserved between the two species, although conjugative DNA transfer has not been observed in *M. tuberculosis*, and ESX-1 does not confer virulence to *M. smegmatis*. The significance of these observations remains to be established.

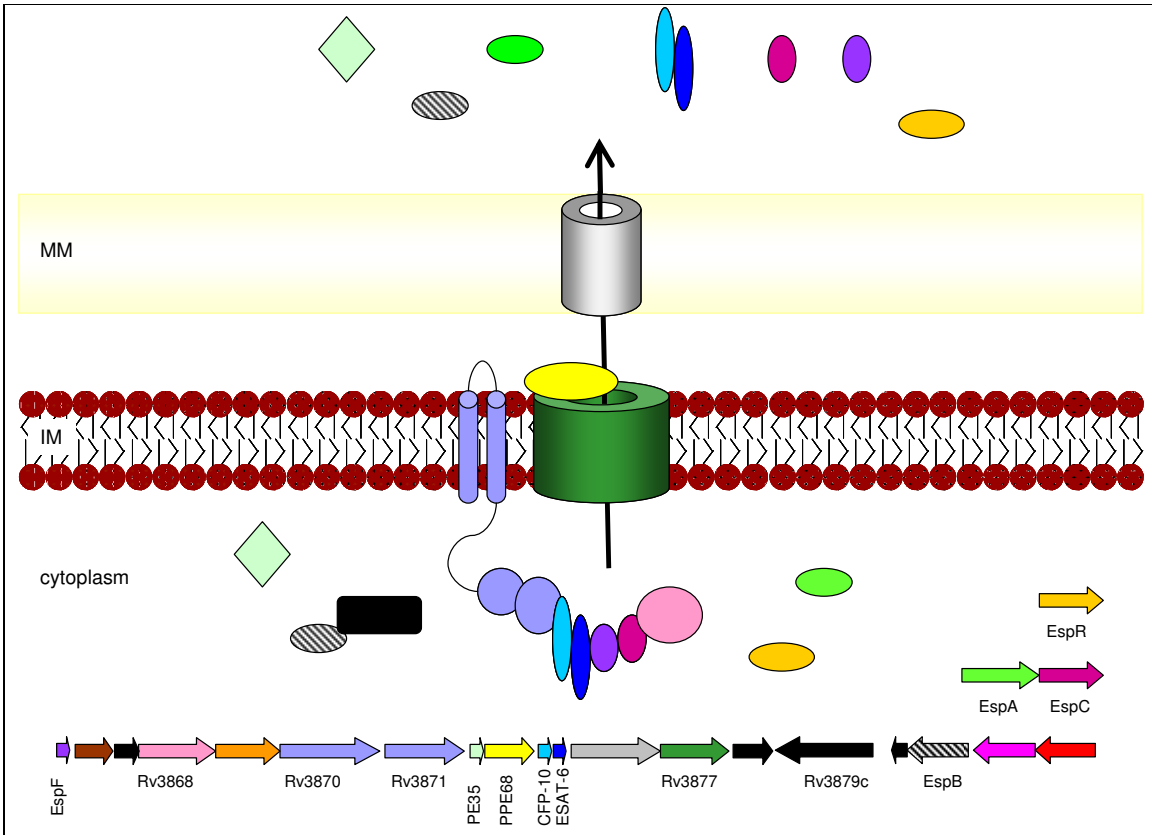


Figure 1.8. The ESX-1 secretion machinery. Various studies have resulted in a model of ESX-1 secretion. ESAT-6 and CFP-10, and EspF and EspC form complexes which interact with each other. CFP-10 and EspC interact with the ATPases Rv3871 and Rv3868 respectively, which provide energy for translocation of the protein complex through the membrane pore, Rv3877. Rv3871 is associated with the membrane via its interaction with Rv3870. EspB interacts with Rv3879c, which may in turn interact with Rv3871 resulting in its secretion. EspA may be associated with the secreted protein complex via Rv3868. All these substrates are dependant on one another for secretion. PPE68 may be a gating protein regulating the secretion of the ESX-1 substrates. PE35 and EspR are also secreted by this secretion machinery via an unknown mechanism. PPE68 and PE35 also interact to form a protein complex, this complex may separate after secretion, releasing PE35, while PPE68 obstructs the pore. The protein responsible for the transport of the substrates through the MM remains unknown. IM is the inner membrane, MM is the mycomembrane.

ESX-2

The ESX-2 secretion system of *M. tuberculosis* is encoded by the genomic region Rv3895c to Rv3884c, located directly adjacent to the ESAT-6 gene cluster region 1 (Gey van Pittius *et al.*, 2001). This secretion system has not been investigated and its function(s) remain unknown.

ESX-3

The ESX-3 secretion system is the only ESX system which is essential for *in vitro* growth of *M. tuberculosis* (Sassetti *et al*., 2003), although ESX-3 is not required for the growth of *M. smegmatis*. Expression of ESX-3 is regulated by divalent cation levels, particularly iron and zinc, as part of the IdeR and Fur/Zur regulons (Rodriguez *et al*., 2002; Maciag *et al*., 2007; Siegrist *et al.,* 2009). Recently Serafini *et al* (2009) have constructed an *M. tuberculosis* ESX-3 conditional mutant in which ESX-3 transcription can be downregulated. They showed that ESX-3 is essential for *M. tuberculosis* survival, but that the mutant can be complemented by high concentrations of iron, zinc or wild-type culture supernatant (Serafini *et al*., 2009). This confirms the role of ESX-3 in iron and zinc homeostasis and suggests that this region is involved in the uptake of divalent cations, possibly by secreting soluble cation-binding proteins. *M. leprae* is unable to produce any siderophores, but is still able to survive and infect, suggesting that it has other mechanisms of iron uptake (Quadri, 2008). As *M. leprae* possesses the ESX-3 secretion system, Serafini *et al* (2009) suggest that ESX-3 may be responsible for iron uptake in this organism. In contrast, Siegrist *et al* (2009) showed that ESX-3 is required for the acquisition of iron from mycobactin, suggesting that *M. leprae* utilises an alternate mechanism of iron acquisition. The ESX-3 of *M. smegmatis* is also regulated by iron concentration, but not by zinc concentration (Maciag *et al*., 2009). Maciag *et al* (2009) suggest that that while iron is limiting in both the human host and the soil environments; *M. smegmatis* is unlikely to experience zinc deficiency in its natural environment, resulting in this difference in regulation.

The role of ESX-3 in iron and zinc homeostasis suggests that ESX-3 may be highly expressed during, and play an important role in the infective process, during which the host restricts the amount of iron available to the pathogen. Siegrist *et al* (2009) have recently shown that ESX-3 is required for growth of *M. tuberculosis* in macrophages. This would also explain the potent antigenicity of EsxH (TB10.4) which has been identified in short-term culture filtrates (Skjot *et al*., 2000). ESX-3 is involved in metal cation uptake, enabling the acquisition of metal ions from mycobactin and possibly zinc transporters. The mechanism by which it functions, its structure and its substrates remain unclear.

ESX-4

The ESAT-6 gene cluster region 4 is the most ancient ESAT-6 gene cluster, from which the other regions were duplicated (Gey van Pittius *et al*., 2001). It is the smallest of the ESAT-6 gene cluster

regions and does not encode any PE or PPE proteins. The ESX-4 secretion system probably performs the original functions of the ESX system, although this remains to be investigated.

Until recently ESX-4 was the only ESX system identified outside of the genus *Mycobacteria*. ESX-4-like clusters have been identified in *Nocardia farcinica*, *Gordonia bronchialis, Corynebacterium diptheriae and* various *Rhodococcus* species, indicating that this secretion system may be conserved amongst the high GC Gram-positive bacteria (N.C. Gey van Pittius, personal communication). Recently another ESX cluster was identified in *N. farcinica* which contains all the conserved components of the larger ESX systems (Bitter *et al* 2009). In addition Esx-like proteins have been identified in various other low GC Gram-positive bacteria including *Bacillus anthracis and Staphylococcus aureus* (Pallen, 2002; Burts *et al.,* 2005; Garufi *et al.,* 2008). Although the only additional member of the ESX system encoded at these loci is the FtsK/SpoIIIE-like protein, the Esx proteins are actively secreted by a mechanism which requires these FtsK/SpoIIE proteins, and secretion thereof is important for virulence.

The ESX-4 secretion machinery is widely spread amongst high GC Gram-positive bacteria and appears to have its roots in a shared ancestor of the low GC Gram-positive bacteria, from which it has expanded and evolved into the immunopathologically important ESX secretion systems present in *M. tuberculosis*.

ESX-5

The ESAT-6 gene cluster region 5 is the most recent duplication, and is found only in the slow-growing mycobacteria (Gey van Pittius *et al*., 2001). There are 3 copies of the PPE and 2 copies of the PE genes in this gene cluster of *M. tuberculosis*, and it appears that these genes have a greater propensity for duplication when associated with this system. Gey van Pittius *et al* (2006) showed that these gene families, especially the PE-PGRS and PPE-MPTR subfamilies, expanded out of the ESX-5 duplication. The *M. marinum* ESX-5 secretion system, in addition to secreting the Esx proteins encoded in this region, also secretes several PPE and PE_PGRS proteins (Abdallah *et al*., 2006; Abdallah *et al*., 2009). ESX-5 may be responsible for the secretion of all the PPE-MPTR and PE-PGRS proteins, which evolved subsequent to the ESX-5 duplication (Gey van Pittius *et al.,* 2006). These ESX-5 substrates are either secreted proteins, or cell surface proteins which appear to be involved in the modulation of the macrophage response against *M. marinum*, mediated by the

expression of cell surface antigens, a reduction in the cytokine response and induction of cell death (Abdallah *et al.*, 2008). Therefore it appears that ESX-5 plays an important role in immune evasion by mycobacteria. The absence of ESX-5, and the PPE-MPTR and PE-PGRS proteins, in fast-growing mycobacteria, which seldom cause disease, supports this role of ESX-5 in immune evasion.

Interactions between the ESX secretions systems

The ESX secretion systems of *M. tuberculosis* appear to be functionally distinct and proteins encoded by one ESAT-6 gene cluster region do not complement homologous proteins knocked out of another region. Interestingly, although regulation of expression is different for each of the ESAT-6 gene clusters, deleting one ESX secretion system can influence the expression of components of the other ESX systems (Abdallah *et al.,* 2006; 2007). Recently, Callahan *et al* (2009) identified protein interactions between EsxA and EsxB (the ESAT-6 and CFP-10 proteins from ESX-1) and Rv3884c and Rv3885c, two proteins encoded in region 2. Whether these ESX-2 proteins can complement their equivalents in ESX-1, and whether they will interact in the presence of the ESX-1 homologs, is unclear. Therefore the nature of the relationship between the different ESX secretion systems remains to be determined.

Conclusion

There are five ESAT-6 gene cluster regions in *M. tuberculosis*, each of which encodes a dedicated ESX secretion system responsible for the secretion of the associated Esx proteins and other substrates. Although some of these secretion systems have been implicated in the virulence of *M. tuberculosis* and ESX-3 is essential for its survival, the specific functions of the secretion systems have not been determined. In addition, the virulence-associated ESX-1 secretion system is present in some non-pathogenic mycobacteria such as *M. smegmatis,* suggesting that this region has a function besides virulence. ESX-3, though essential in *M. tuberculosis*, is not essential (and is differentially regulated) in *M. smegmatis.* Although several substrates of the various secretion systems have been identified, hinting at possible functions of the systems, the precise and distinct role of each ESX secretion system remains undeciphered. In addition, it remains unclear as to how these ESX secretion systems interact with one another. Despite extensive research on these virulence-associated ESX secretion systems, very little is clearly understood regarding their substrates, structure, mechanism, regulation and functions; elucidation of which may lead to developments in the prevention and treatment of tuberculosis disease.

CHAPTER 2

The *Mycobacterium tuberculosis* ESX-3 secretion system interactome

2.1. Introduction

The ESAT-6 gene cluster region 3 is the only ESAT-6 gene cluster that is essential for growth in *M. tuberculosis* (Sassetti *et al.,* 2003), but it is not essential in the fast-growing saprophyte, *M. smegmatis*. Expression of ESX-3 is regulated by iron and zinc availability as part of the IdeR and Fur/Zur regulons (Rodriguez *et al*., 2002; Maciag *et al*., 2007), and may be involved in divalent cation homeostasis. Recently Serafini *et al* (2009) have constructed a conditional ESX-3 knockout of *M. tuberculosis*, in which down regulation of ESX-3 expression is lethal to the organism, but may be complemented by iron, zinc or wild-type culture filtrate (Serafini *et al*., 2009). It is clear that the ESX-3 secretion system is involved in metal cation homeostasis, and may be involved in the secretion of factors which are involved in the uptake of these, and other, metal cations.

Three additional Esx proteins, EsxQ (Rv3017c), EsxR (Rv3019c), and EsxS (Rv3020c), located outside of the ESAT-6 gene clusters in the *M. tuberculosis* genome, were shown to be duplicated from the ESAT-6 gene cluster region 3 (Gey van Pittius *et al*., 2006). The amino acid sequences of EsxG and EsxS are 91% identical, while EsxQ and EsxR are 65% and 84% identical to EsxH, respectively. These Esx proteins are also secreted by *M. tuberculosis*, and are regulated by Zur in response to zinc concentration (Maciag *et al*., 2007), although they do not appear to be repressed by IdeR.

This study aims to elucidate the structure of the ESX-3 secretion system, encoded by the ESAT-6 gene cluster region 3, by identifying interacting proteins encoded in the gene cluster and to investigate interactions between EsxG, H, Q, R and S, the Esx proteins encoded in, and duplicated from, the ESAT-6 gene cluster region 3. Elucidating the structure of the ESX-3 secretion system will serve as a model for the structure of the ESX secretion systems, advance our understanding of the mechanisms of mycobacterial pathogenicity and may assist in providing clues as to how these secretion mechanisms may be interfered with, potentially leading to developments in the prevention and treatment of tuberculosis disease.

2.2. Experimental approach

This study makes use of a mycobacterial-2-hybrid system called Mycobacterial – Protein Fragment Complementation (M-PFC), developed and described by Singh *et al.* (2006), to identify interacting proteins between components of the ESX-3 secretion system. This two-hybrid technique uses a heterologous mycobacterial host organism, *M. smegmatis,* which is fast growing and non-pathogenic, and allows for the identification of interacting mycobacterial proteins. M-PFC has the advantage over traditional yeast-2-hybrid systems for the identification of mycobacterial interacting proteins in that the intracellular environment of *M. smegmatis* is more similar to that of *M. tuberculosis* and as such, allows for the correct protein expression, folding, modifications and interactions which may not occur in an unrelated host.

The system is based on the functional reconstitution of two murine dihydrofolate reductase (mDHFR) domains fused to the proteins of interest, expressed from the M-PFC expression vectors, pUAB300 and pUAB400, in *M. smegmatis*. mDHFR reconstitution confers resistance to the antibiotic trimethoprim, allowing for the selection of clones containing interacting proteins on media containing the antibiotic (Singh *et al.,* 2006). The M-PFC technique, as used in this study, is outlined in Figure 2.1.

The ESAT-6 gene cluster region 3 consists of 11 genes, which were cloned into each of the M-PFC expression vectors in a total of 12 fragments, due to size constraints. In addition, the EsxG-EsxH operon was also cloned into pUAB300 to allow coexpression of EsxG and EsxH together with each of the other ESX-3 proteins. Each combination of vectors was cotransformed into *M. smegmatis* to allow for the detection of interacting protein pairs. EsxQ, EsxR and EsxS were also cloned into each vector and cotransformed with EsxG, EsxH and each other to identify interactions between the ESX-3-associated Esx proteins.

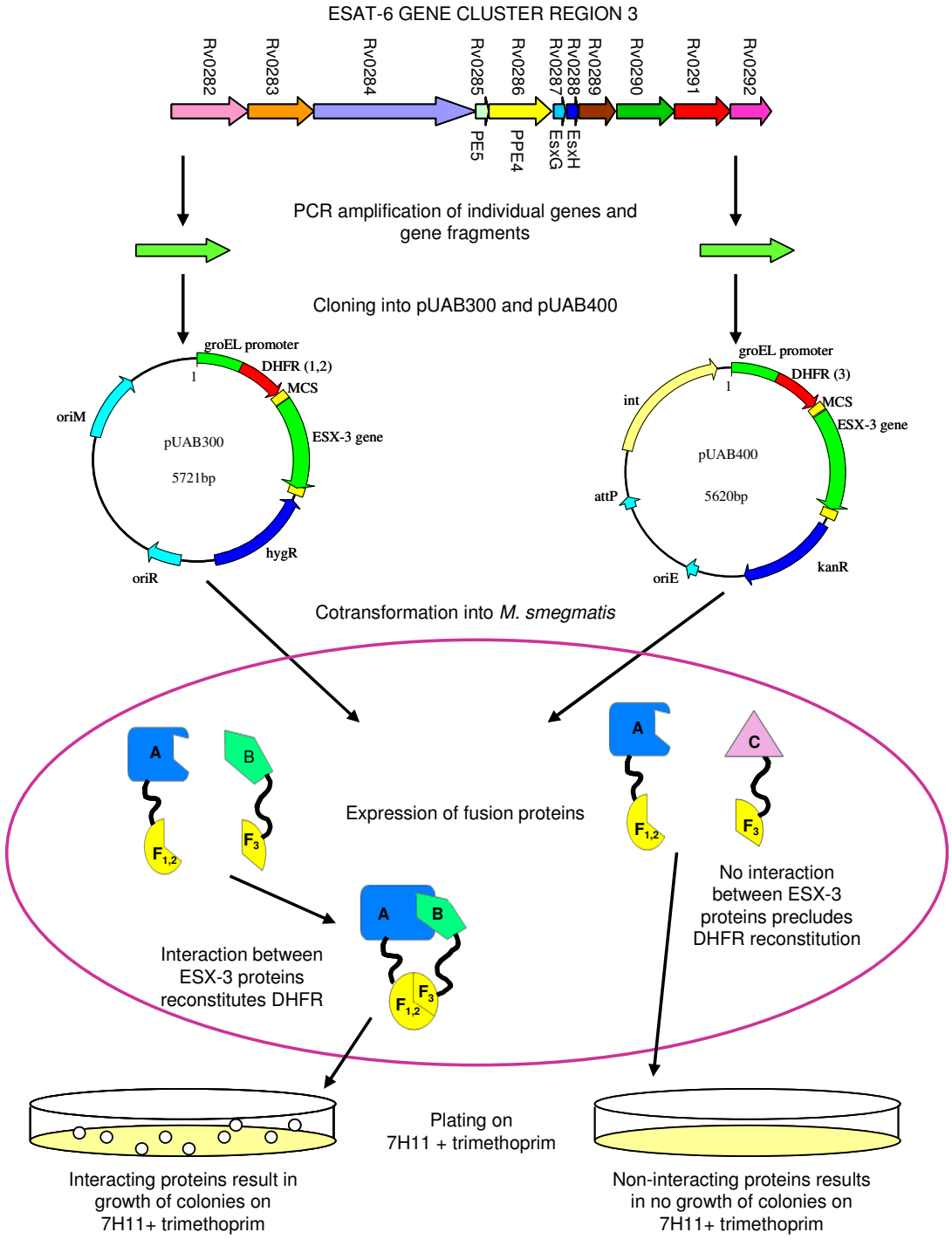


Figure 2.1. Identification of interacting proteins of ESX-3 using M-PFC. Each gene from the *M. tuberculosis* ESAT-6 gene cluster region 3 was PCR amplified and cloned in frame into each of the M-PFC expression vectors, pUAB300 and pUAB400. Each pair of expression vectors were cotransformed into *M. smegmatis*. The DHFR-ESX-3 fusion proteins were expressed. Should the two ESX-3 proteins interact, the DHFR is reconstituted and is able to break down the antibiotic trimethoprim, resulting in the growth of colonies on trimethoprim-containing solid medium. If the two ESX-3 proteins do not interact, the trimethoprim becomes toxic to the *M. smegmatis* cells and no colonies form on trimethoprim-containing medium.

2.3. Materials and Methods

All standard molecular techniques were performed essentially as described by Sambrook *et al.* (1989). Techniques and reagents are described in detail in Appendix A.

2.3.1. Bacterial strains

E. coli strain JM109 was used for standard cloning procedures and *E. coli* K12 ER2925 (New England Biolabs (NEB)) for methylase-sensitive cloning. *M. smegmatis* mc^2155 was used as a heterologous mycobacterial host organism for the detection of protein-protein interactions.

2.3.2. Growth Media and Culture conditions

E. coli was cultured in LB with shaking, and on LB agar plates, overnight at 37 ℃. Solid and liquid media were supplemented with antibiotics; ampicillin (50 ug/ml, Roche), kanamycin (50 ug/ml, Sigma) and hygromycin (100 ug/ml, Invitrogen); as appropriate.

M. smegmatis was grown in Middlebrook 7H9 media (BD) with shaking, and on BBL™ Seven H11 Agar Base (BD) plates, at 37 ℃ for 2-3 days or as otherwise stated. Solid and liquid media were supplemented with 0.5% glycerol, 0.5% glucose and 0.2% Tween-80. Kanamycin (25 ug/ml, Sigma), hygromycin (50 ug/ml, Invitrogen) and trimethoprim (15-50 ug/ml, Sigma) were added as appropriate.

2.3.3. PCR primers

All DNA and protein sequence information was obtained from the *M. tuberculosis* H37Rv genome sequence database (Tuberculist, http://genolist.pasteur.fr/TubercuList/). Primers were designed using Primer Premier 5.0 (PREMIER Biosoft International) and obtained from IDT (through Whitehead Scientific, www.whitesci.co.za). Primers were designed to amplify all 12 genes and gene fragments from ESX-3 such that the resultant PCR products contain ClaI restriction sites on each end, and could be cloned in frame into the expression vectors, pUAB300 and pUAB400 (Table 2.1).

Table 2.1. PCR primers used for the amplification of the ESX-3 genes for cloning into the M-PFC vectors

Name of primer	Primer sequence (from 5' to 3')	Primer length	Tm [°C]	Gene/ gene fragment amplified	Product length (bp)
Rv0282 pUAB400 f	ATCGATATGGCGGGCGTAGGTGAA	24	62.1	Rv0282 in frame for plasmid vectors pUAB400 and pUAB300	1929
Rv0282 pUAB300 f	ATCGATGATGGCGGGCGTAGGTGAA	25	64.0		1930
Rv0282 pUAB300/400 r	ATCGATCGTGGTCGTGCTGCTGGTT	25	64.4		
Rv0283 pUAB400 f	ATCGATGCGGGCATGACGAACCA	23	63.3	Rv0283 in frame for plasmid vectors pUAB400 and pUAB300	1663
Rv0283 pUAB300 f	ATCGATTGCGGGCATGACGAACCA	24	63.2		1664
Rv0283 pUAB300/400 r	ATCGATCGGCGACGAGCCTCAAAGA	25	63.7		
Rv0284a pUAB400 f	ATCGATGTGAGCAGACTGATCTTTG	25	56.9	Rv0284 up to base 2007 in frame for plasmid vectors pUAB400 and pUAB300	2036
Rv0284a pUAB300 f	ATCGATGGTGAGCAGACTGATCTTT	25	58.0		2037
Rv0284a pUAB300/400 r	ATCGATGACTCGATGTGGTAGGC	23	58.5		
Rv0284b pUAB400 f	ATCGATAGCGTTTCTCGCCAGATC	24	58.7	Rv0284 from base 1973 in frame for plasmid vectors pUAB400 and pUAB300	2030
Rv0284b pUAB300 f	ATCGATCAGCGTTTCTCGCCAGATC	25	60.8		2031
Rv0284b pUAB300/400 r	ATCGATCATGACTGACTCCCCTTCTG	26	59.7		
Rv0285 pUAB400 f	ATCGATATGACGTTGCGAGTGGTTCCG	27	62.2	Rv0285 in frame for plasmid vectors pUAB400 and pUAB300	324
Rv0285 pUAB300 f	ATCGATCATGACGTTGCGAGTGGTTCC	27	62.4		325
Rv0285 pUAB300/400 r	ATCGATTGCTCAGCCGCCCACGAC	24	65.6		
Rv0286 pUAB400 f	ATCGATTACGTACGGGGTCGTGG	23	60.5	Rv0286 in frame for plasmid vectors pUAB400 and pUAB300	1592
Rv0286 pUAB300 f	ATCGATCTACGTACGGGGTCGTG	23	60.0		1593
Rv0286 pUAB300/400 r	ATCGATTCGGAATTCGGTTACTTGCT	26	58.5		
Rv0287 pUAB400 f	ATCGATGTTATGAGCCTTTTGGATGCTCA	29	60.1	Rv0287 in frame for plasmid vectors pUAB400 and pUAB300	300
Rv0287 pUAB300 f	ATCGATTATGAGCCTTTTGGATGCTCA	27	58.4		301
Rv0287 pUAB300/400 r	ATCGATCAGAACCCGGTATAGGTCGACGC	29	64.0		
Rv0288 pUAB400 f	ATCGATGTGATGTCGCAAATCATGTACA	28	58.6	Rv0288 in frame for plasmid vectors pUAB400 and pUAB300	388
Rv0288 pUAB300 f	ATCGATTGTGATGTCGCAAATCATGTACA	29	58.7		389
Rv0288 pUAB300/400 r	ATCGATCGCCCCAATGGTTTCAGC	24	61.9		
Rv0289 pUAB400 f	ATCGATATGGATGCAACACCCAACGC	26	61.4	Rv0289 in frame for plasmid vectors pUAB400 and pUAB300	1000
Rv0289 pUAB300 f	ATCGATCATGGATGCAACACCCAACG	26	61.4		1001
Rv0289 pUAB300/400 r	ATCGATCCTGCTGTCCGCAAGAATGG	26	62.7		
Rv0290 pUAB400 f	ATCGATCCAGGAACGGTCCCGCTGAT	26	65.2	Rv0290 in frame for plasmid vectors pUAB400 and pUAB300	1468
Rv0290 pUAB300 f	ATCGATCCCAGGAACGGTCCCGCTGAT	27	66.6		1469
Rv0290 pUAB300/400 r	ATCGATCACGCAAATGCGGCACGGA	25	65.1		
Rv0291 pUAB400 f	ATCGATATTGTTCGCTTGGGTGCTC	25	59.3	Rv0291 in frame for plasmid vectors pUAB400 and pUAB300	1514
Rv0291 pUAB300 f	ATCGATGATTGTTCGCTTGGGTGCTC	26	61.3		1515
Rv0291 pUAB300/400 r	ATCGATTGTCGATTGCCAGGGGTAG	25	60.6		
Rv0292 pUAB400 f	ATCGATGGAGCCCACCGAATGAAC	24	61.2	Rv0292 in frame for plasmid vectors pUAB400 and pUAB300	1058
Rv0292 pUAB300 f	ATCGATGGGAGCCCACCGAATGAAC	25	62.9		1059
Rv0292 pUAB300/400 r	ATCGATGGACTATCTGCGGCGGATGA	26	60.6		

Name of primer	Primer sequence (from 5' to 3')	Primer length	Tm [°C]	Gene/ gene fragment amplified	Product length (bp)
EsxQ pUAB400 f	GGAGATCGATGTGGTGTCGCAGAGTATG	28	60.4	EsxQ in frame for plasmid vectors pUAB400 and pUAB300	385
EsxQ pUAB300 f	ATCGATCGTGGTGTCGCAGAGTATG	25	60.3		382
EsxQ pUAB300/400 r	ATCGATCGTTTAGTCTCCGGCGTC	24	60.5		
EsxR pUAB400 f	GAGAATCGATGTGATGTCGCAGATTATG	28	57.4	EsxR in frame for plasmid vectors pUAB400 and pUAB300	212
EsxR pUAB300 f	ATCGATTGTGATGTCGCAGATTATG	25	55.4		209
EsxR pUAB300/400 r	ATCGATGTGCTAGCCGCCCC	20	62.4		
EsxS pUAB400 f	GAGAATCGATGTGATGAGTTTGTTGGATG	29	57.8	EsxS in frame for plasmid vectors pUAB400 and pUAB300	315
EsxS pUAB300 f	ATCGATTGTGATGAGTTTGTTGGATG	26	55.8		312
EsxS pUAB300/400 r	ATCGATTTTTAAAACCCGGTGTAGC	25	56.0		
F102	AGAACCACCACGAGGAGCTCAT	22	60.3	Sequencing primers for pUAB300 insert confirmation	
R102	TGATGCCTGGCAGTCGATCGTA	22	60.4		
pUAB400 F	TCCCAGAATACCCAGGCGTCCTCT	25	63.5	Sequencing primers for pUAB400 insert confirmation	
pUAB400 R	AGGCCCAGTCTTTCGACTGAGCCTTTC	27	63.9		

2.3.4. PCR amplification

The *M. tuberculosis* H37Rv genomic DNA was used as a template for the PCR amplification of the gene components of the ESAT-6 gene cluster region 3. The ESX-3 genes, and EsxQ, EsxR and EsxS, were amplified using FastStart Taq DNA Polymerase (Roche) as per manufacturers instructions, using the GC-rich buffer and annealing temperatures (Tm) as described in Table 2.1. Elongation times were calculated at 1 minute per 1kb. The products were separated by agarose gel electrophoresis in a 1.5% agarose gel in TAE, and visualised using UV detection. The appropriate bands were excised and the DNA purified using the Wizard® SV PCR and Gel Clean-up kit (Promega).

2.3.5. Cloning vectors

The desired inserts were cloned into pGemT-Easy (Promega) and subcloned into the ClaI restriction sites in the *E. coli* – mycobacterial shuttle vectors pUAB300 and pUAB400 (Kind gift from A. Steyn, University of Alabama at Birmingham). pUAB300 is an episomal vector, encoding resistance for hygromycin, while pUAB400 is an integrating vector which confers kanamycin resistance. pUAB300 and pUAB400 are mycobacterial expression vectors which produce protein fusions of the inserted DNA and the mDHFR fragments 1 and 2, and 3 respectively. Interaction of the fusion proteins expressed from the vectors allows functional reconstitution of mDHFR and confers resistance to trimethoprim. Vectors used in this study are described in Table 2.2.

2.3.6. Cloning into pGemT-Easy

The PCR products were ligated into the pGemT-Easy vector (Promega) and transformed into *E. coli* K12 ER2925 cells (NEB). Transformants were selected on LB agar plates with ampicillin (50 ug/ml) and PCR screened to verify the presence of the insert. Plasmids were extracted from 10 ml cultures using the Wizard® *Plus* SV Miniprep Plasmid Purification Kit (Promega), quantified and the constructs sequenced to confirm the sequence of the insert.

Table 2.2. Vectors used in the identification of protein-protein interactions in ESX-3

Vector/construct		Description	Size (bp)	Source/Reference
pGEM-T easy		*E. coli* cloning T-vector, Amp^R, *lacZ*, *ori*E	3015	Promega
pUAB400		*E. coli*-Mycobacterial shuttle vector, integrative, Kan^R, groEL promoter, DHFR (3), attP, oriE, int	4720	Singh *et al.* 2006
pUAB300		*E. coli*-Mycobacterial shuttle vector, episomal, Hyg^R, groEL promoter, DHFR (1,2),	4913	Singh *et al.* 2006
pUAB400_	Rv0282	pUAB400 expressing DHFR(3) fused to Rv0282	6649	This study
pUAB400_	Rv0283	pUAB400 expressing DHFR(3) fused to Rv0283	6383	This study
pUAB400_	Rv0284a	pUAB400 expressing DHFR(3) fused to Rv0284b	6756	This study
pUAB400_	Rv0284b	pUAB400 expressing DHFR(3) fused to Rv0284a	6750	This study
pUAB400_	Rv0285	pUAB400 expressing DHFR(3) fused to Rv0285 (PE5)	5044	This study
pUAB400_	Rv0286	pUAB400 expressing DHFR(3) fused to Rv0286 (PPE4)	6312	This study
pUAB400_	Rv0287	pUAB400 expressing DHFR(3) fused to Rv0287 (EsxG)	5020	This study
pUAB400_	Rv0288	pUAB400 expressing DHFR(3) fused to Rv0288 (EsxH)	5108	This study
pUAB400_	Rv0289	pUAB400 expressing DHFR(3) fused to Rv0289	5720	This study
pUAB400_	Rv0290	pUAB400 expressing DHFR(3) fused to Rv0290	6188	This study
pUAB400_	Rv0291	pUAB400 expressing DHFR(3) fused to Rv0291 (MycP3)	6234	This study
pUAB400_	Rv0292	pUAB400 expressing DHFR(3) fused to Rv0292	5778	This study
pUAB400_	EsxQ	pUAB400 expressing DHFR(3) fused to EsxQ (Rv3017c)	5105	This study
pUAB400_	EsxR	pUAB400 expressing DHFR(3) fused to EsxR (Rv3019c)	4932	This study
pUAB400_	EsxS	pUAB400 expressing DHFR(3) fused to EsxS (Rv3020c)	5035	This study
pUAB300_	Rv0282	pUAB300 expressing DHFR(1,2) fused to Rv0282	6843	This study
pUAB300_	Rv0283	pUAB300 expressing DHFR(1,2) fused to Rv0283	6577	This study
pUAB300_	Rv0284	pUAB300 expressing DHFR(1,2) fused to Rv0284b	6950	This study
pUAB300_	Rv0284	pUAB300 expressing DHFR(1,2) fused to Rv0284a	6944	This study
pUAB300_	Rv0285	pUAB300 expressing DHFR(1,2) fused to Rv0285 (PE5)	5238	This study
pUAB300_	Rv0286	pUAB300 expressing DHFR(1,2) fused to Rv0286 (PPE4)	6506	This study
pUAB300_	Rv0287	pUAB300 expressing DHFR(1,2) fused to Rv0287 (EsxG)	5214	This study
pUAB300_	Rv0288	pUAB300 expressing DHFR(1,2) fused to Rv0288 (EsxH)	5302	This study
pUAB300_	Rv0289	pUAB300 expressing DHFR(1,2) fused to Rv0289	5914	This study
pUAB300_	Rv0290	pUAB300 expressing DHFR(1,2) fused to Rv0290	6382	This study
pUAB300_	Rv0291	pUAB300 expressing DHFR(1,2) fused to Rv0291(MycP3)	6428	This study
pUAB300_	Rv0292	pUAB300 expressing DHFR(1,2) fused to Rv0292	5972	This study
pUAB300_	EsxQ	pUAB300 expressing DHFR(1,2) fused to EsxQ (Rv3017c)	5295	This study
pUAB300_	EsxR	pUAB300 expressing DHFR(1,2) fused to EsxR (Rv3019c)	5122	This study
pUAB300_	EsxS	pUAB300 expressing DHFR(1,2) fused to EsxS (Rv3020c)	5225	This study
pUAB300	Rv0287-88	pUAB300 expressing DHFR(1,2) fused to Rv0287; and Rv0288 (DHFR(1,2)-EsxG, EsxH)	5632	This study

Amp^R – ampicillin resistance; Kan^R – kanamycin resistance; Hyg^R – hygromycin resistance; *ori*E – *E. coli* origin or replication; *ori*M – Mycobacterial origin of replication; DHFR – dihydrofolate reductase fragment; attP – attachment site of phage; int – integrase

2.3.7. Subcloning into pUAB300 and pUAB400

The pGem constructs and pUAB300 and pUAB400 vectors were digested with ClaI (Promega). The digests were separated electrophoretically and the inserts and digested expression vectors were excised and purified. The linearized pUAB300 and pUAB400 vectors were dephosphorylated using Shrimp Alkaline Phosphatase (SAP, Roche). The inserts were ligated into the vectors using T4 DNA ligase (Promega) and transformed into competent *E. coli* JM109 cells. Transformants were selected on LB plates with hygromycin (100 ug/ml, pUAB300) or kanamycin (50 ug/ml, pUAB400). The size and orientation of each insert was verified by PCR amplification using the forward sequencing primer and the insert-specific reverse PCR primer. The plasmids were purified and quantified and the sequence and orientation of the inserts were confirmed by sequencing. The M-PFC constructs generated in this study are described in Table 2.2.

2.3.8. Transformation of the pUAB400 (pUAB400:ESX-3) constructs into *M. smegmatis*

The purified pUAB400 constructs containing the ESX-3 genes were transformed into *M. smegmatis* mc^2155 cells and transformants selected for on 7H11 medium containing kanamycin (25 ug/ml). Transformants contain the pUAB400 construct integrated into the *M. smegmatis* genome. Colonies were picked and PCR screened to confirm the presence of the insert.

2.3.9. Cotransformation of pUAB300 constructs (pUAB300:ESX-3) into *M. smegmatis*

Electrocompetent cells were prepared for each pUAB400 construct-containing *M. smegmatis* clone. Each pUAB300 construct was transformed individually into each pUAB400 construct-containing *M. smegmatis* clone. Transformants were selected for on 7H11 medium containing kanamycin (25 ug/ml) and hygromycin (50 ug/ml). Colonies were screened to confirm the presence of the pUAB300 constructs.

2.3.10. Identification of interacting protein pairs.

M. smegmatis clones containing both a pUAB300:ESX-3 and a pUAB400:ESX-3 construct were plated on 7H11 medium containing kanamycin (25 ug/ml), hygromycin (50 ug/ml) and trimethoprim (15-50 ug/ml), and incubated at 37 ℃ for 10-12 days. Clones which grew on trimethoprim contained proteins which

interact, allowing functional reconstitution of mDHFR and conferring trimethoprim resistance. The *M. smegmatis* clones which were positive for growth on trimethoprim contain interacting proteins. Colonies from each of the positive clones were resuspended in 7H9 and replated on 7H11 with kanamycin (25 ug/ml), hygromycin (50 ug/ml) and trimethoprim (15-50 ug/ml) and incubated for 7 days at 37 ℃ to confirm the growth.

2.3.11. Positive and negative controls

Rv0287 (EsxG) and Rv0288 (EsxH) have previously been shown to interact (Okkels and Andersen, 2004); therefore *M. smegmatis* containing the pUAB300_Rv0287 and pUAB400-Rv0288 vectors was used as a positive control. *M. smegmatis* containing pUAB300_Rv0287 and the empty pUAB400 vector was used as a negative control.

2.3.12. Graph Editor

Protein interaction networks were generated using YED graph editor software (Tübingen, Germany, www.yWorks.com).

2.3.13. Protein structure analysis

The structure and functional domains of the ESX-3 protein components were predicted by *in silico* analysis using protein sequence data from Tuberculist, the TMHMM Server v2.0 transmembrane prediction server (TMHMM, http://www.cbs.dtu.dk/services/TMHMM/), SignalP 3.0 signal peptide prediction server (SignalP, http://www.cbs.dtu.dk/services/SignalP/) and the Conserved Domain Database (CDD, http://ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). These protein analyses were used in the construction of the model of the ESX-3 secretion machinery.

2.4. Results

2.4.1. M-PFC – ESX-3 construct cloning

The gene components of ESX-3, as well as the genes for EsxQ, EsxR and EsxS, were successfully amplified, cloned into pGem-T Easy, and subcloned, in frame, into the M-PFC vectors pUAB300 and pUAB400. The sequence and orientation of the inserts were confirmed by sequencing. The integrating pUAB400 constructs were transformed into *M. smegmatis,* and verified by PCR. Competent cultures were prepared from single colonies and confirmed by ZN staining. Each episomal pUAB300 construct was individually transformed into each pUAB400 construct-containing *M. smegmatis* and the construct presence was confirmed by PCR.

2.4.2. Optimisation of trimethoprim concentration for selection of interacting proteins

Initial screening for growth of *M. smegmatis* containing interacting proteins on media containing 30 ug/ml and 50 ug/ml trimethoprim identified only an interaction between Rv0287 (EsxG) and Rv0288 (EsxH). Protein interaction between EsxG and EsxH has been previously observed (Okkels and Andersen, 2004). The Esx protein pairs have been shown to form tight complexes, a fact which together with their small size (~100 aa) may allow high levels of DHFR reconstitution, thereby conferring resistance to high trimethoprim concentrations. Many of the other ESX-3 proteins are larger and membrane-associated and likely bind less strongly than the EsxG-EsxH dimer; or only interact transiently. We therefore screened for interactions between the EsxG, H, Q, R and S proteins, together with the negative control (*M. smegmatis* containing pUAB300_Rv0287 and pUAB400), on trimethoprim concentrations ranging from 15 to 30 ug/ml to identify a trimethoprim concentration which consistently allowed for the identification of interacting protein pairs, but did not allow growth of the negative control. We expected to see interactions beween the Esx proteins pairs EsxG-EsxH and EsxR-EsxS, and possibly EsxQ-EsxS. In addition, homodimerisation of the Esx proteins has been previously observed. Trimethoprim concentrations of 20 ug/ml and 25 ug/ml both allowed for the identification of various interacting Esx proteins, however only screening on 15 ug/ml consistently identified protein interactors in both plasmid combinations; with no growth of the negative control within 12-14 days. Therefore all future screenings for interactions were done using 15 ug/ml trimethoprim concentration. The use of 15 ug/ml trimethoprim for the identification of interacting proteins is

supported by other groups using the M-PFC technique (D. Crossman, S. Sampson, A. Steyn; personal communication).

2.4.3. Interactions between the ESX-3-associated Esx proteins

The Esx proteins are secreted substrates of the ESX secretion systems. We investigated protein-protein interactions between the Esx proteins encoded in ESX-3, and duplicated from it. The M-PFC constructs containing the ESX-3-associated Esx genes, EsxG, H, Q, R and S, were transformed into *M. smegmatis* in all possible combinations. Clones containing interacting protein pairs were selected on medium containing 15 ug/ml trimethoprim. Interactions were detected between the EsxG-EsxH and EsxR-EsxS protein pairs. In addition we observed interactions between the other ESX-3 associated ESAT-6/CFP-10 pairs EsxG-EsxQ, EsxG-EsxR, EsxH-EsxS and EsxQ-EsxS, in both construct combinations. We also observed the homodimerisation of each of the Esx proteins examined; as well as interactions between the ESAT-6 homologs (EsxH, EsxQ and EsxR) and the CFP-10 homologs (EsxG and EsxS). Colonies were observed within 4 days for *M. smegmatis* clones containing an ESAT-6 and CFP-10 homolog, while colonies were only observed after 4 days for those clones containing two CFP-10- or ESAT-6-like proteins. In essence, all of the ESX-3 associated Esx proteins interacted with themselves and one another (Figure 2.2), although stronger growth was observed for those *M. smegmatis* clones which expressed both an ESAT-6 and a CFP-10 homolog.

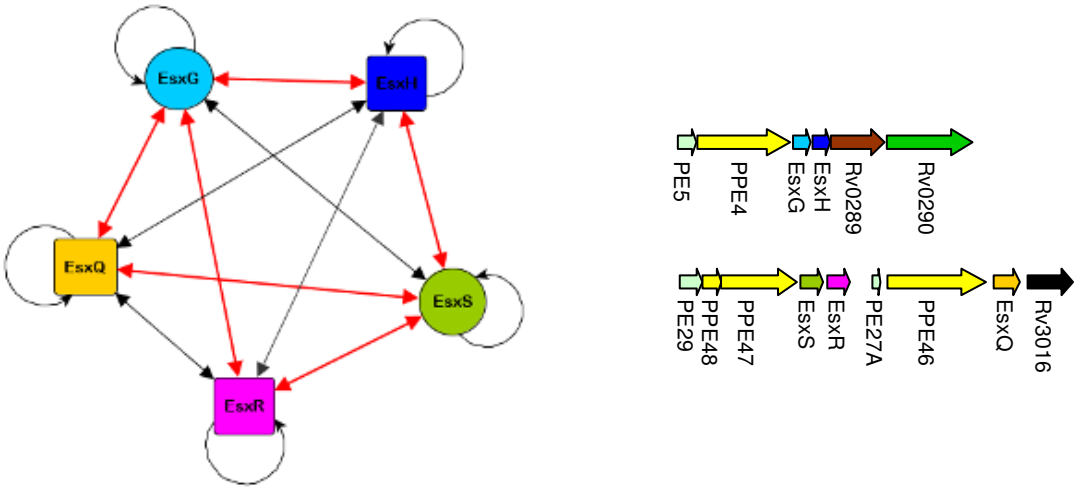


Figure 2.2. The ESX-3-associated Esx proteins interact to form hetero- and homo-dimers. The ESAT-6 gene cluster region 3-encoded Esx proteins, EsxG and EsxH, and the Esx proteins duplicated from region 3, EsxQ, EsxR and EsxS, interact with themselves and with each other. Stronger growth was observed in *M. smegmatis* containing a CFP-10 and ESAT-6 homolog than between two CFP-10 or ESAT-6 homologs. EsxG and EsxH are encoded within the ESAT-6 gene cluster region 3; while EsxQ, EsxR and EsxS are encoded in a cluster of *pe, ppe* and *esx* genes, all duplicated from region 3. The Esx proteins represented by squares are ESAT-6 homologs, circles represent CFP-10 homologs. Red arrows indicate strong interactions, black arrows indicate weaker interactions.

2.4.4. The ESX-3 interactome

In order to construct an interactome of the ESX-3 secretion system, we used M-PFC to identify protein-protein interactions between all combinations of the ESX-3 proteins. The pUAB400:ESX-3 and pUAB300:ESX-3 constructs were cotransformed into *M. smegmatis* in all possible combinations, and clones containing interacting protein pairs were selected on medium containing 15 ug/ml trimethoprim (Table 2.3). Each combination of proteins was represented in both vector combinations, and those pairs of proteins which resulted in growth of *M. smegmatis* in both vector combinations were taken as positive interactions. In addition, proteins which were shown to interact with themselves (represented in a single vector combination), were regarded as positive. Clones expressing both EsxG and EsxH and another ESX-3 protein were considered to be positive when either of the Esx proteins was also found to interact with the protein of interest. The growth of these *M. smegmatis* clones containing interacting proteins was confirmed by replating on trimethoprim-containing medium. Figure 2.3 is a network of the observed ESX-3 protein-protein interactions. Specific interactions and their proposed roles are described in the discussion.

Table 2.3. Interacting ESX-3 proteins identified by M-PFC

pUAB300 construct	pUAB400 construct: Rv0282	Rv0283	Rv0284a	Rv0284b	Rv0285	Rv0286	Rv0287	Rv0288	Rv0289	Rv0290	Rv0291	Rv0292
Rv0282	X	X	**X**		**X**	X	X	**X**	**X**		X	**X**
Rv0283		X	**X**						**X**			
Rv0284a	**X**	**X**	X	**X**	**X**	**X**	**X**	X	**X**		X	**X**
Rv0284b			**X**		**X**			X		**X**	X	X
Rv0285	**X**	X	**X**	**X**	X			X	**X**	**X**	X	**X**
Rv0286		X	**X**	X			X	**X**				**X**
Rv0287			**X**	X			X	**X**			X	
Rv0288	**X**					**X**	**X**	X	**X**		X	**X**
Rv0289	**X**	**X**	**X**		**X**		X	**X**			X	
Rv0290		X	X	**X**	**X**					X	X	X
Rv0291												
Rv0292	**X**	X	**X**		**X**	**X**	X	**X**	X			
Rv0287-88		X	X	X			X	X				

Protein combinations for which growth was observed on 15 ug/ml trimethoprim are indicated with an X. Bold **X** indicated that growth was observed for *M. smegmatis* containing the genes in both vector combinations. Protein combinations which are shaded in grey were considered to be positive interactions.

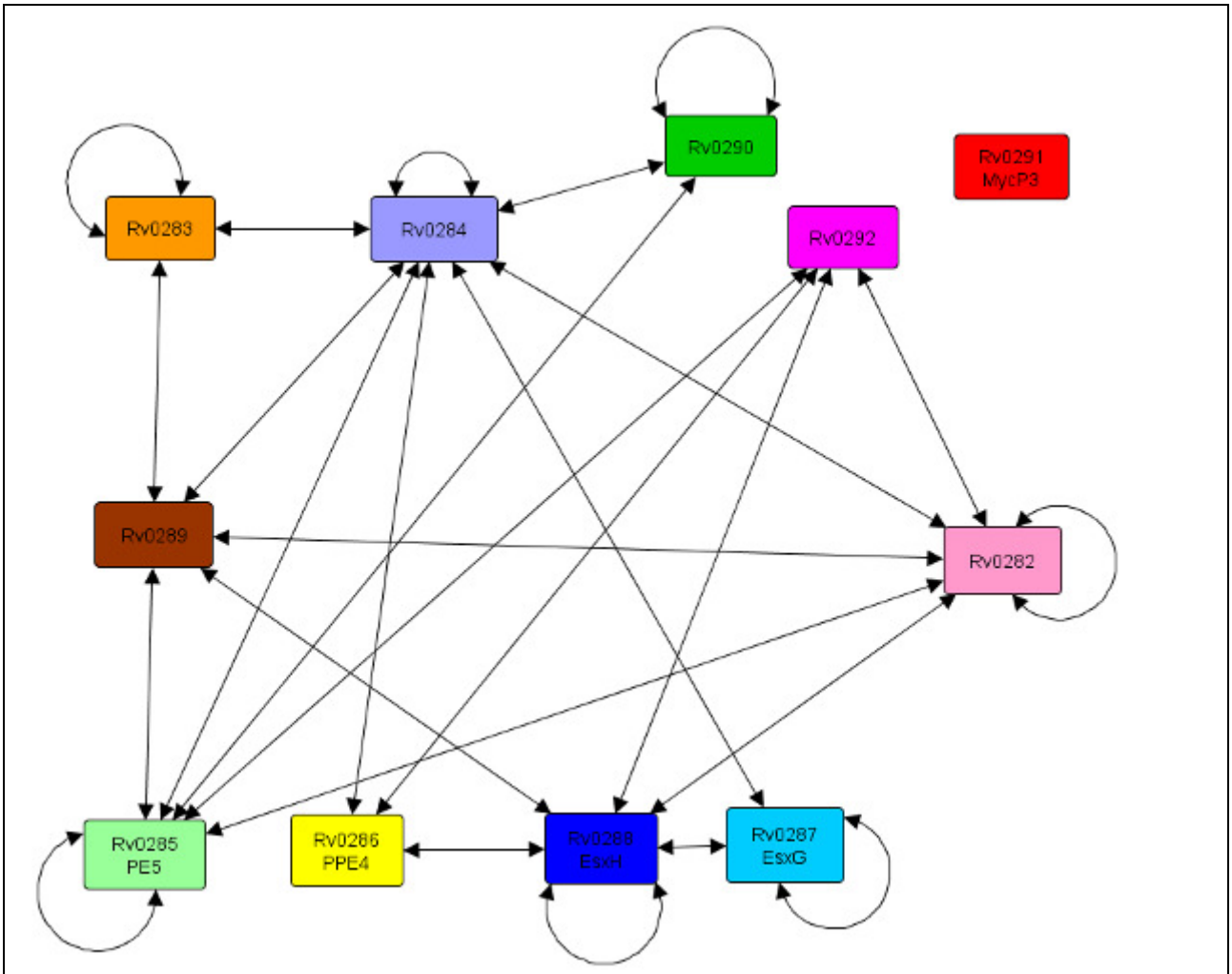


Figure 2.3. The ESX-3 interactome of protein-protein interactions identified in this study. The *M. tuberculosis* ESX-3 proteins form a complex web of interactions. The proteins are represented in 3 tiers, the top tier representing the putative membrane proteins, the lowest the putative ESX-3 substrates and the remaining cytoplasmic proteins in the middle. For simplicity, the Rv0284 is represented as a single unit combining the interactions of Rv0284a and b.

2.4.5. ESX-3 *in silico* protein structure predictions

The structure and functional domains of each of the *M. tuberculosis* ESX-3 secretion system proteins, encoded by Rv0282 to Rv0292, were analysed using data from Tuberculist, TMHMM, SignalP and CDD. The results obtained are are consistent with those of Bitter *et al.* (2009). The properties of these proteins were used in constructing a model of the ESX-3 secretion system, based on the protein-protein interactions identified and other previously determined interactions and functions of ESX proteins.

Rv0282

Rv0282 is a conserved hypothetical protein of 631 amino acids, which shows homology to SpoVK and cbbX-cfxQ family proteins. It contains an AAA+ class ATPase domain of the P-loop NTPase superfamily in the region between amino acids 360 and 420. This domain includes Walker A and B motifs, an arginine finger and an ATP binding site. ATPases of this family are chaperone-like ATPases involved in the assembly and disassembly of protein complexes. No transmembrane helices or signal sequences were identified in this protein. Rv0282 appears to be a cytoplasmic ATPase involved in the assembly of protein complexes required for ESX-3 secretion.

Rv0283

Rv0283 is a possible conserved membrane protein of 538 amino acids. This protein contains a hydrophobic transmembrane domain in the N-terminus (Figure 2.4)

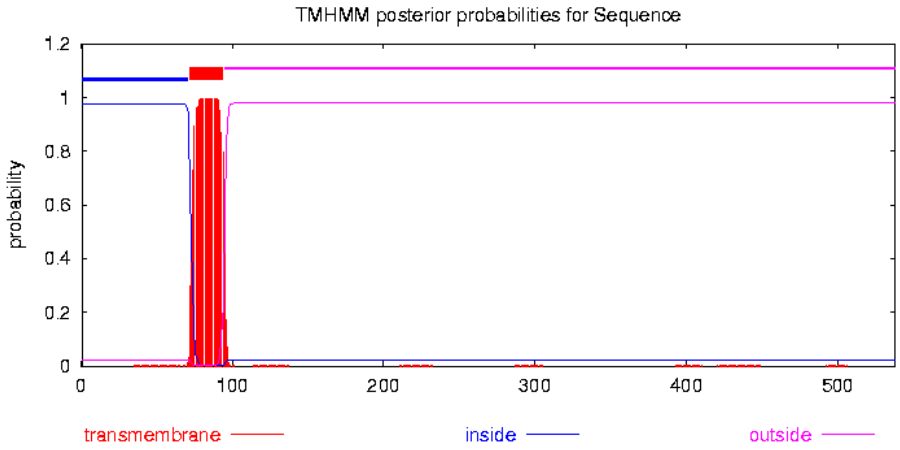


Figure 2.4. Rv0283 is an N-terminal membrane protein. The TMHMM profile of Rv0283 indicates that this protein contains an N-terminal transmembrane helix.

Rv0284

Rv0284 is a conserved membrane protein of 1330 amino acids which contains one (possibly two) N-terminal transmembrane domain (Figure 2.5) and two cytoplasmic C-terminal ATPases of the P-loop NTPase superfamily. Rv0284 shows homology to FtsK-SpoIIIE proteins, which are predicted to be involved in the translocation of plasmid or chromosomal DNA during conjugative DNA transfer and septum formation (Wu *et al.,* 1994; Begg *et al.,* 1995) and may be responsible for providing the energy for translocation of the ESX-3 secreted substrates.

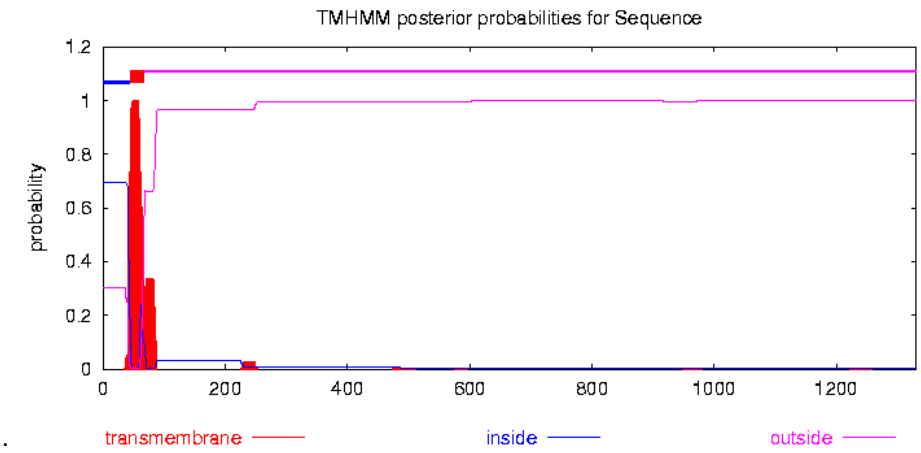


Figure 2.5. Rv0284 contains N-terminal transmembrane domains. The TMHMM profile of Rv0284 identifies an N-terminal transmembrane helix. The TMHMM profiles of homologs of Rv0284 suggest that the lower peak adjacent to the transmembrane peak may be a second transmembrane helix.

Rv0285 (PE5)

Rv0285 is a PE family protein of 102 amino acids, which contains a putative N-terminal signal peptide with a probable cleavage site between amino acid 31 an 32, at the sequence ASA-AP. The PE proteins are putative secreted substrates of the ESX secretion systems.

Rv0286 (PPE4)

Rv0286 is a member of the PPE family of proteins. The protein is 513 amino acids in length and contains a PPE superfamily domain in the N-terminal 160 amino acids. TMHMM analysis of this protein identified 3 transmembrane domains in the centre of the protein, between amino acids 240 and 340 (Figure 2.6). This is not congruent with the PPE proteins from the other ESAT-6 gene clusters (Figure 2.7). The PPE proteins encoded in the *M. tuberculosis* ESAT-6 gene clusters 1, 2 and 5 contain no transmembrane motifs, or contain N-terminal signal sequence-like motifs, and are predicted to be secreted by the ESX secretion systems. The structure of the PPE4 homolog in *M. smegmatis* is, however, similar to that of *M. tuberculosis*, also containing three transmembrane domains, suggesting that this structure is essential for the functioning of this PPE protein in mycobacteria.

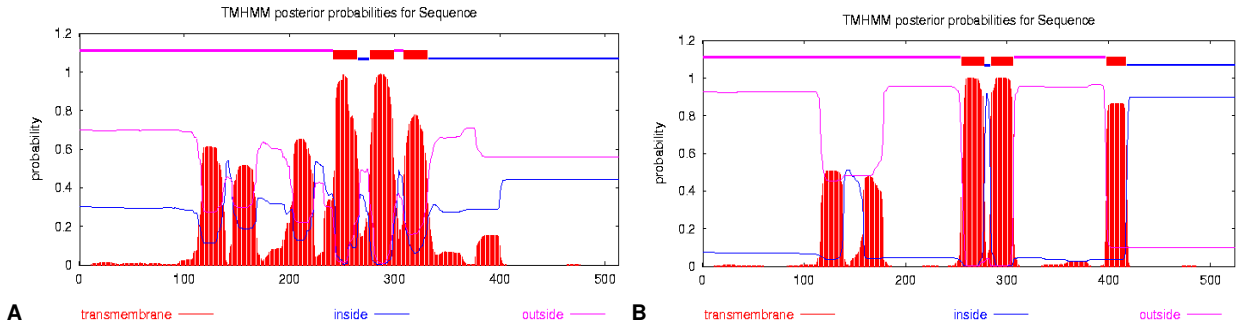


Figure 2.6. Rv0286 (PPE4) contains 3 transmembrane-like domains in the middle of the peptide. The PPE proteins encoded by the ESAT-6 gene cluster region 3 of *M. tuberculosis* (**A**) and *M. smegmatis* (**B**) both contain 3 transmembrane domains.

To further investigate the unusual structure of PPE4 generated by TMHMM, we compared the TMHMM probability of PPE4 with the PPE proteins that are duplicated from and associated with PPE4 (Figure 2.8, Gey van Pittius *et al* 2006). Similar TMHMM profiles were observed for the PPE proteins which were duplicated directly from PPE4, with the features becoming less prominent in the ESX-3 associated PPEs. The ESAT-6 gene cluster region 3 is the first duplication of the ESAT-6 gene cluster 4, and the first region to contain a PPE protein (N.C. Gey van Pittius, personal communication). It appears, therefore, that the structure of the PPE proteins has evolved significantly subsequent to the duplications from PPE4.

The PPE proteins contain a conserved N-terminal domain, while their C- terminal regions are highly variable (Cole *et al*., 1998; Camus *et al*., 2002). PPE proteins often occur in an operon with a PE protein, with which it interacts (Tundup *et al*., 2006; Tundup *et al*., 2008), and the complex of which is secreted by its associated ESX secretion system (Abdallah *et al*., 2006; Abdallah *et al*., 2009). PPE4 occurs within an operon with PE5, in the ESAT-6 gene cluster region 3, and contains the conserved N-terminal PPE domain; and the hydrophobic transmembrane-like domains occurring in the more variable C-terminus. It appears that the properties of the PPE proteins which result in their secretion by the ESX secretion systems are derived from PPE4, while the hydrophobic transmembrane-like domains have been lost in subsequent duplications. Therefore we propose that PPE4 is secreted via the ESX-3 secretion machinery and subsequently integrated into the mycomembrane via its 3 transmembrane domains.

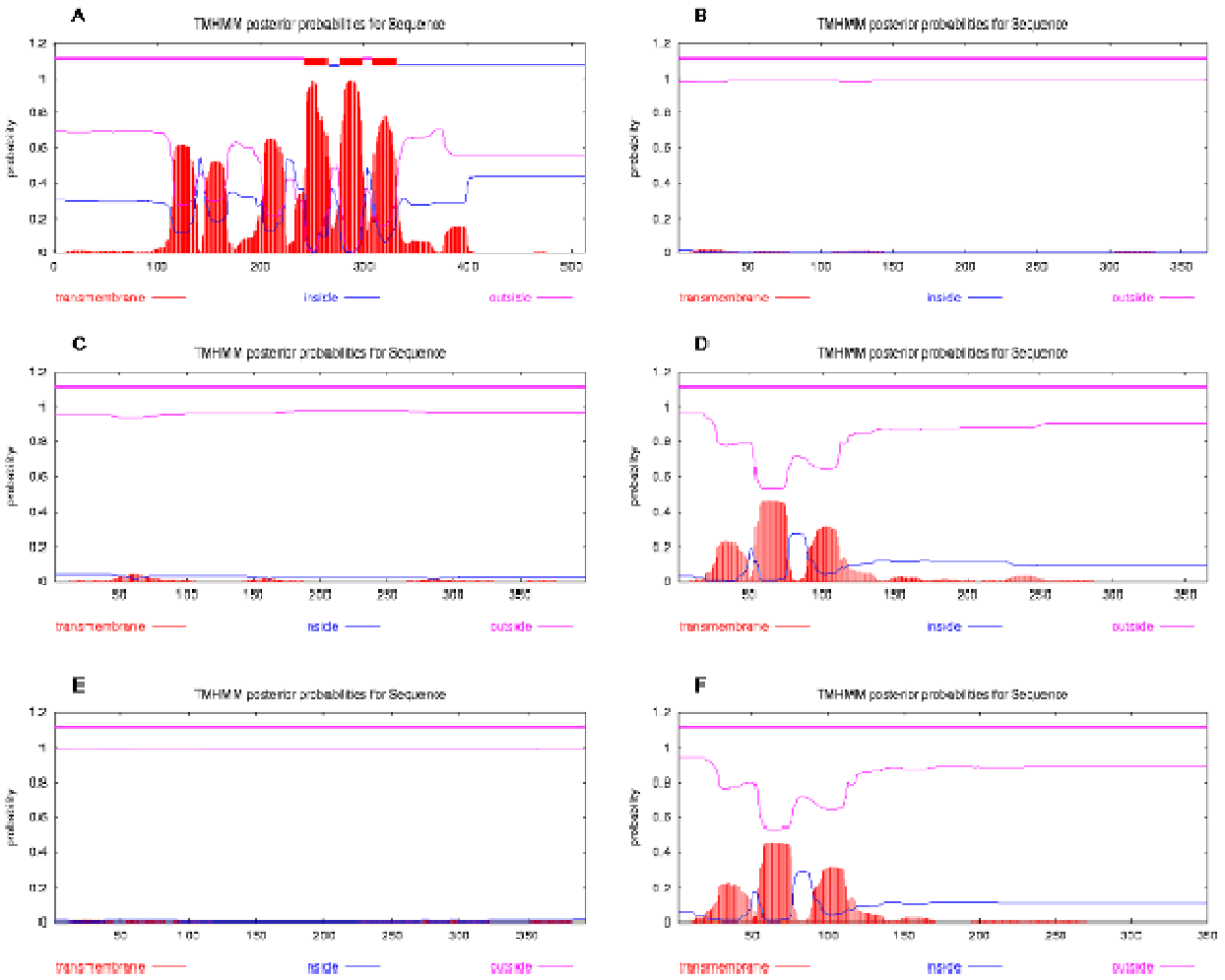


Figure 2.7. The transmembrane structure of PPE4 is unique among the ESAT-6 gene cluster-encoded PPE proteins. The TMHMM profile of PPE4, encoded by ESX-3 (A), differs considerably from those of the PPE proteins encoded by ESX-1 (PPE68, B) and ESX-2 (PPE69, C) which contain no hydrophobic transmembrane-like motifs, and the ESX-5-encoded PPEs, PPE25 (D), PPE26 (E) and PPE27 (F). Signal sequence-like motifs are observed in the N-terminals of PPE25 and PPE27, but no further transmembrane-like motifs are present.

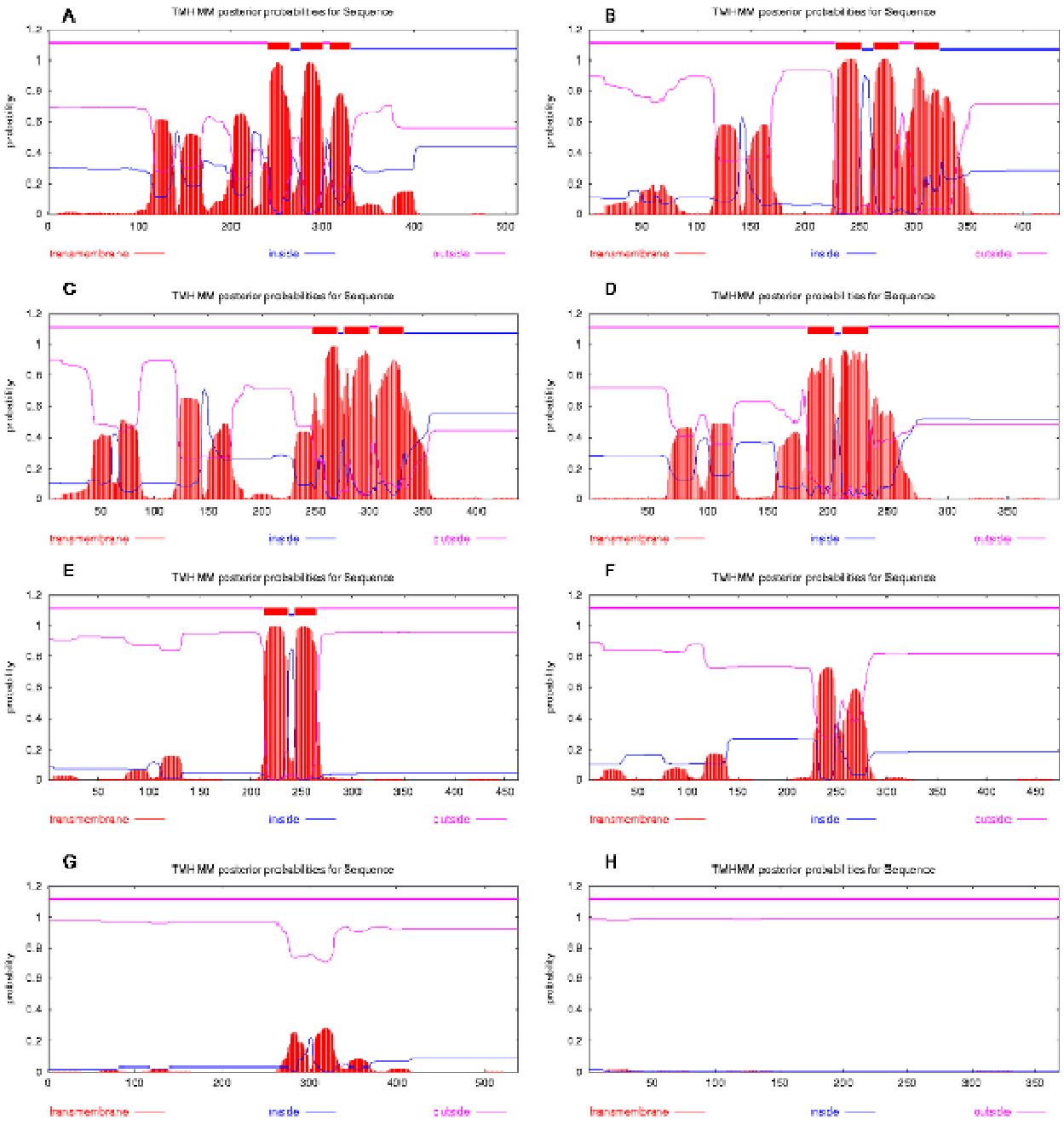


Figure 2.8. Transmembrane-like motifs are present in several PPE proteins which are duplicated from, or associated with PPE4. The PPE proteins which were duplicated directly from PPE4 (A); PPE46 (B) and PPE47-48 (C) exhibit TMHMM profiles similar to PPE4, maintaining the 3 mid-protein transmembrane domains and 2 hydrophobic stretches further upstream. Additional signal sequence-like motifs begin to emerge in the N-terminus. It appears that the PPE proteins have evolved away from this structure as although the PPE4-associated protein PPE66-67 (D) exhibits similar structure, these properties are less prominent in other PPE proteins associated with PPE4, PPE1 (E), PPE37 (F), PPE20 (G) and are absent from the ESX-1-encoded PPE68 (H).

Rv0287 (EsxG)

Rv0287 is a CFP-10-like Esx protein of 97 amino acids, a secreted substrate of ESX-3.

Rv0288 (EsxH)

Rv0288 is a 96 amino acid ESAT-6-like Esx protein, a potent antigen (TB10.4) secreted by ESX-3.

Rv0289

Rv0289 is a conserved hypothetical protein of 295 amino acids, which has a signal sequence-like motif in the N-terminus (Figure 2.9), and may therefore be a secreted protein.

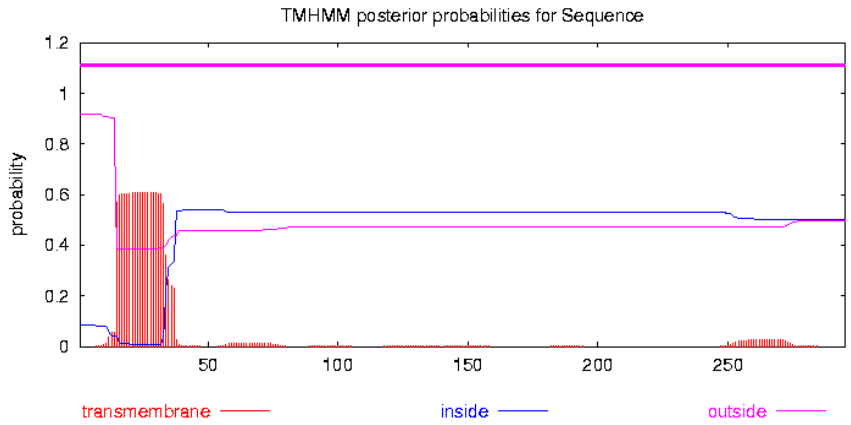


Figure 2.9. The N-terminus of Rv0289 contains a signal sequence-like motif.

Rv0290

Rv0290 is a conserved integral membrane protein of 472 amino acids which consists of 11 transmembrane helices (Figure 2.10) and is predicted to be the membrane pore through which the ESX-3 substrates are transported through the cell membrane (Tekaia *et al.*, 1999; Gey van Pittius *et al.*, 2001).

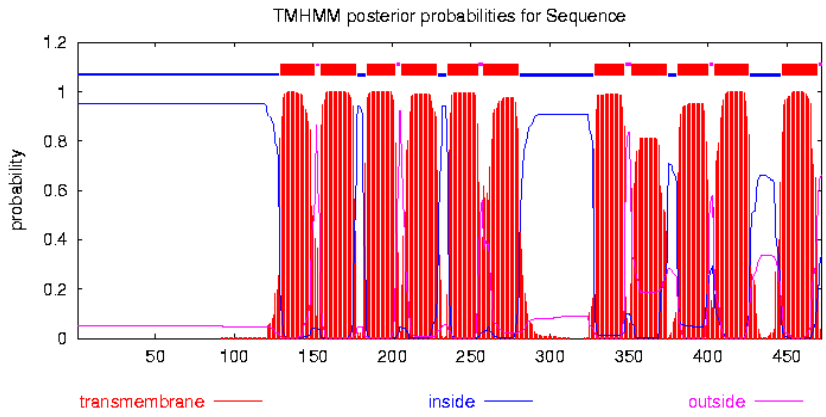


Figure 2.10. Rv0290 contains 11 transmembrane domains.

Rv0291

Rv0291 is a 461 amino acid membrane anchored, cell wall-associated mycosin (subtilisin-like serine protease), MycP3. MycP3 contains an N-terminal signal-sequence motif and a C-terminal transmembrane helix (Figure 2.11), suggesting that most of this protein occurs outside of the cell and is anchored to the cell membrane via the C-terminus. The N-terminal is likely cleaved between amino acids 25 and 26 (AWA-IG), after secretion. This protein contains a peptidase-like motif of the subtilisin superfamily between amino acids 75 and 375,

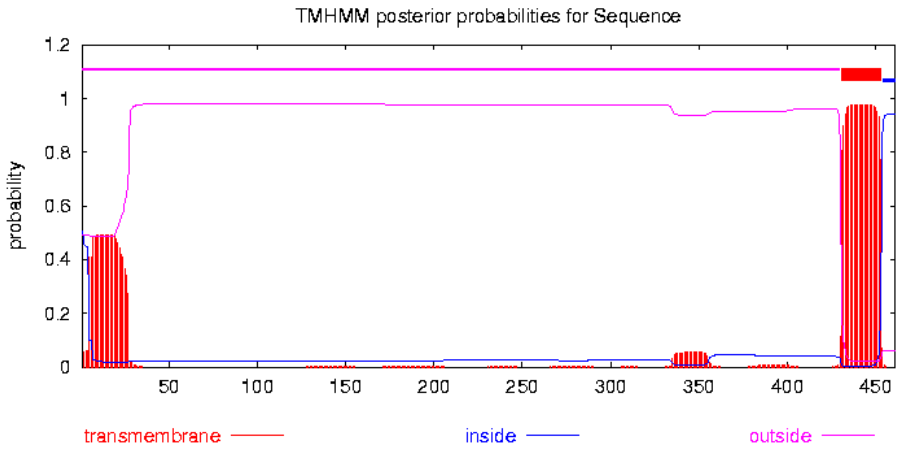


Figure 2.11. Rv0291 is a secreted, membrane anchored mycosin protease. The TMHMM profile clearly depicts the N-terminal signal sequence and C-terminal transmembrane domain.

Rv0292

Rv0292 is a conserved transmembrane protein of 331 amino acids, which contains two N-terminal transmembrane domains (Figure 2.12) of which one might be a signal peptide, as predicted by SignalP, and is likely cleaved between amino acids 26 an d 27.

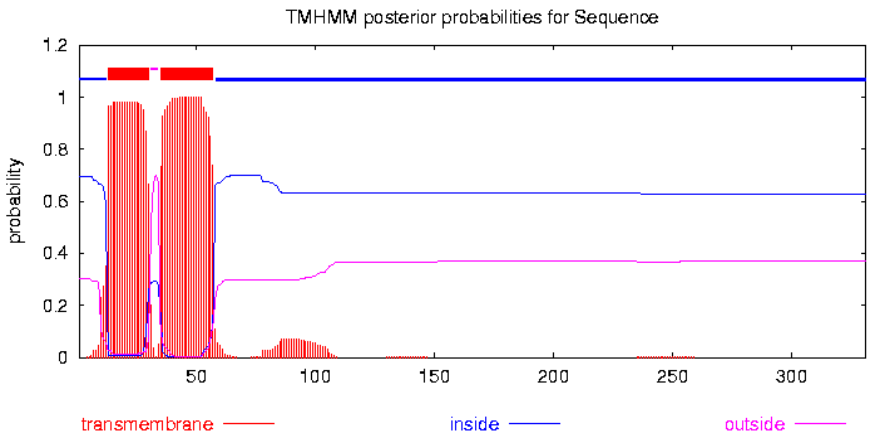


Figure 2.12. Rv0292 is a transmembrane protein. Two N-terminal transmembrane domains are predicted by TMHMM analyses although one of these may be a signal peptide, as predicted by SignalP 3.0.

2.5. Discussion

The ESX-3-associated Esx proteins form hetero- and homodimers

The M-PFC screen of the Esx proteins EsxG, EsxH, EsxQ, EsxR and EsxS indicates that all of these proteins can interact as ESAT-6/CFP-10 pairs, as well as ESAT-6/ESAT-6 and CFP-10/CFP-10 pairs, of identical and related proteins. This is consistent with the results of other studies which have shown interactions between the Esx protein pairs, and detected self-dimerisation of individual Esx proteins (Stanley *et al.*, 2003; Okkels and Andersen, 2004; Singh *et al.*, 2006; Teutschbein *et al.*, 2009).

It has not previously been shown that the Esx proteins interact with other Esx proteins outside of their protein pairs, Okkels *et al.* (2004) specifically showed that EsxG and EsxH do not interact with EsxA and EsxB (ESX-1 Esx proteins). However, the EsxQ, EsxR and EsxS proteins are duplicated from EsxG and EsxH, and are fairly homologous to these proteins (65-84% identical vs 21-29% identity between EsxG/EsxH and EsxB/A), which could explain the ability of these proteins to interact. Interestingly, all five ESX-3-associated Esx proteins are under transcriptional regulation by Zur (zinc dependant), together with ESX-3 (Maciag *et al.*, 2007). The conservation of the regulation of EsxQ, EsxR and EsxS by Zur subsequent to their duplication suggests that their function remains linked to ESX-3. ESX-3 is also transcriptionally regulated by IdeR in response to iron (Rodriguez *et al.*, 2002), however only EsxG and EsxH are under control of IdeR, suggesting that these Esx proteins may be differentially expressed and function under different environmental conditions.

The physiological importance of the homodimerisation of the Esx proteins remains to be established, as these interactions have only been detected in *in vitro* studies using recombinant proteins and in two-hybrid experiments where the protein occurs in the absence of its partner. These proteins may not actually interact when coexpressed with their partners, due to the strength of interaction of the heterodimer. However, heterotetramers consisting of dimers of the ESAT-6/CFP-10 complex have been observed in Western blotting experiments (N.C. Gey van Pittius, personal communication), suggesting that the Esx proteins may form larger complexes, lending credibility to these interactions.

The membrane proteins of ESX-3 interact to form a membrane complex

We investigated the interactions between the membrane components of ESX-3. Three ESX-3 membrane proteins were found to dimerise in this study; Rv0283, Rv0290 and the N-terminal half of Rv0284 (Rv0284a). Rv0284a and Rv0284b, the N- and C- terminal halves of Rv0284, also interacted, which is expected as they form a single protein. In addition Rv0283, Rv0290 and Rv0292 interact with Rv0284 (Figure 2.13).

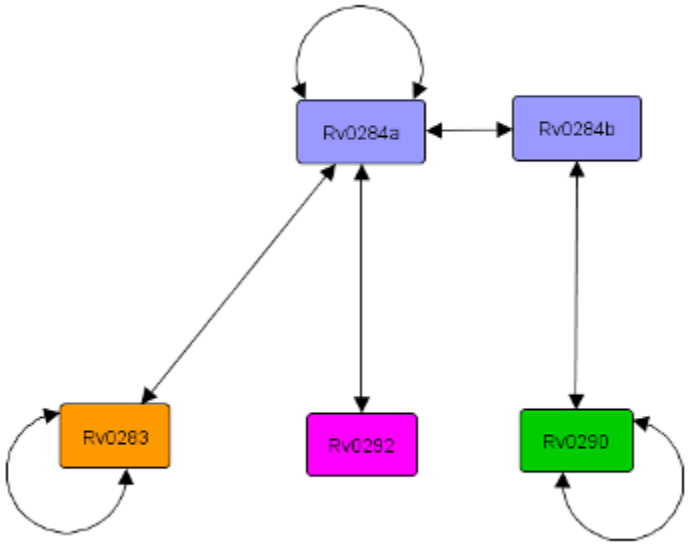


Figure 2.13. The membrane components of ESX-3 interact to form a membrane protein complex.

The interactions between these membrane proteins have not previously been identified in any of the ESX systems, and this may be due to their localisation in the membrane. Rv0283 and Rv0284 possess one and two transmembrane domains respectively, while Rv0292 has two transmembrane domains. Rv0290 consists of 11 transmembrane domains and is predicted to form the membrane pore through which the substrates of ESX-3 are secreted. One of the limitations of the yeast-2-hybrid system is its inability to identify membrane interactions. In addition, the unique mycobacterial membrane structure may be required for proper folding of these proteins or stabilisation of the complex, which would also prevent *in vitro* detection of interactions.

Rv0284 is a membrane-associated ATPase, where the N-teminal part of the protein penetrates the membrane and the C-terminus contains the ATPase domains. Therefore it appears that the membrane section (Rv0284a) interacts with itself and the membrane proteins Rv0283 and Rv0292, while the ATPase domains interact with Rv0290. The C-terminal of the homolog of this protein in ESX-1 interacts with CFP-10 and is predicted to provide the energy for translocation of ESAT-6/CFP-10 through the membrane pore (Renshaw *et al.*, 2002; Hsu *et al.*, 2003; Stanley *et al.*, 2003; Guinn *et al.*, 2004; Renshaw *et al.*, 2005; Champion *et al.*, 2006). Rv0284 presumably binds to Rv0290 to enable the translocation of EsxG-EsxH through the pore structure. The functions of Rv0292 and Rv0283 have not been established, although the ESX-1 homolog of Rv0292 is required for the secretion of ESAT-6 (Brodin *et al.,* 2006) and therefore presumably forms an essential part of the secretion complex. These proteins may be involved in the stabilisation of the membrane complex, or assist in recruiting the substrates to the secretion machinery.

The proteins required for the transfer of secreted ESX substrates through the mycomembrane have not been identified. We have observed dimerisation of Rv0290, suggesting that this protein may form multiple pore structures which may extend through the plasma membrane and mycomembrane to transfer the ESX substrates directly from the cytosol into the extracellular milieu, although this interaction and complex formation would need to be verified.

We propose that the membrane proteins Rv0283, Rv0284, Rv0290 and Rv0292 together form a membrane complex, possibly incorporating multiple copies of some of the proteins, which, powered by the ATPase activity of Rv0284 transfer the ESX substrates directly through the plasma membrane, via a channel formed by Rv0290.

Secreted substrates of ESX-3 interact with a variety of ESX-3 proteins

It has been suggested that in addition to the Esx protein complex, the ESX secretion systems secrete their associated PE-PPE protein complex (Abdallah *et al.*, 2006; Abdallah *et al.*, 2009). Previous studies have detected homodimerisation of both PE and PPE proteins, and well as the formation of PE-PPE protein complexes (Tundup *et al.*, 2006; Teutschbein *et al.*, 2009). The ESAT-6 gene cluster region 3 encodes the PE5-PPE4 protein pair. In this study we detected homodimerisation of PE5 (Rv0285), however we did not detect homodimerisation of PPE4 (Rv0286) or complex formation of PE5-PPE4. Tundup *et al* (2006)

suggest that PE and PPE proteins will form homodimers when expressed on their own, and will only form heterodimers when co-transcribed and -translated, as interactions between the proteins during translation are required for stable complex formation.

Presuming that PE5 and PPE4 do indeed form a secreted complex, we compared the interacting partners of the PE5-PPE4 and EsxG-EsxH complexes. We observed that, with one exception, PE5-PPE4 and EsxG-EsxH exhibit identical interactomes. Both complexes interact with the membrane proteins Rv0284 and Rv0292 and with the cytoplasmic proteins Rv0282 and Rv0289 (Figure 2.14). PE5-PPE4 also interacts with Rv0290, via PE5.

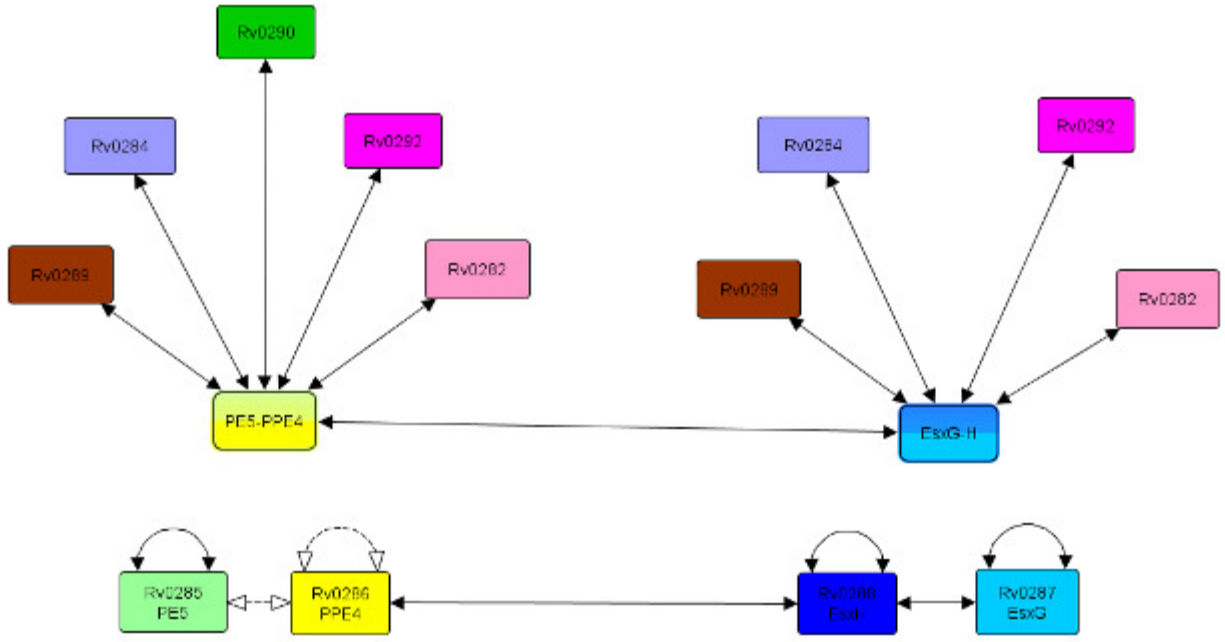


Figure 2.14. The putative substrates of ESX-3 have almost identical interactomes. PE5 and PPE4, and EsxG-EsxH are proposed to form protein complexes. The combined interactions of each protein pair with other proteins encoded by ESX-3 are plotted, and are identical, with the exception of the interaction between PE5-PPE4 and Rv0290. In addition the two protein complexes interact with each other. Solid arrows depict interactions identified in this study, while dotted arrows represent putative interactions which have been shown for related protein pairs.

Rv0284 is the membrane ATPase which is proposed to provide the energy for ESX substrate translocation and is therefore, presumably involved in the secretion of all ESX-3 substrates. Rv0284 binds to both

complexes, and through the hydrolysis of ATP allows the secreted complexes to be transported through Rv0290.

Rv0282 is a putative chaperone protein, which has ATPase activity. Rv0282 may be involved in complex formation by providing the energy required to cause binding of the proteins. The Rv0282 homolog, Rv3868, is required for ESAT-6 secretion (Brodin *et al.,* 2006), and oligomerises to form a hexamer which possesses ATPase activity which results in changes to its structure, and may assist in protein-protein binding (Ogura *et al*., 2004; Luthra *et al*., 2008). Rv0282 also binds to itself, supporting a similar structure and function for this homolog. Rv0282 also binds to Rv0284, Rv0289 and Rv0292, suggesting that it may rather be involved in catalysing the interaction between the substrate complexes and one or more of these proteins.

The only difference between the interactions of EsxG-EsxH and PE5-PPE4 is the interaction between PE5-PPE4 and Rv0290. It has been suggested that the PPE proteins may function as gating proteins (Brodin *et al*., 2006) controlling the secretion of the Esx proteins, as knocking out PPE68 (in ESX-1) causes an increase in ESAT-6 secretion. Rv0290 is predicted to form the membrane pore, and binding of PPE4 to the pore may obstruct EsxG-EsxH secretion. However, an interaction was only detected between PE5 and Rv0290, therefore it is unclear whether this may be the function of PPE4, unless PE5 is an anchoring protein which connects PPE4 to Rv0290, to prevent Esx secretion. The PE-PPE complexes are predicted to be additional ESX system substrates and therefore the reduction in Esx secretion may rather result from competition between the PE-PPE and Esx-Esx complexes for secretion by the ESX system.

Rv0289 is a cytoplasmic protein with an N-terminal signal sequence-like motif and unknown function. Rv0289 interacts with several proteins in addition to PE5-PPE4 and EsxG-EsxH; Rv0282, Rv0283 and Rv0284 (Figure 2.15). The ESX-1 homolog of Rv0289, Rv3866, is not required for ESAT-6 secretion, however its deletion results in an ESAT-6-independant reduction in virulence (Brodin *et al.,* 2006). This, together with the N-terminal signal sequence-like motif, suggests that Rv3866 and Rv0289 may be additional substrates of their respective ESX secretion systems. The only difference between the EsxG-EsxH and Rv0289 interactomes is that Rv0289 binds to Rv0283 and not to Rv0292, which may indicate that different membrane proteins are involved in the secretion of different substrates. However, if Rv0289

is a secreted substrate, and is secreted independently of the Esx and PE/PPE complexes, the role of the interactions between them is unclear.

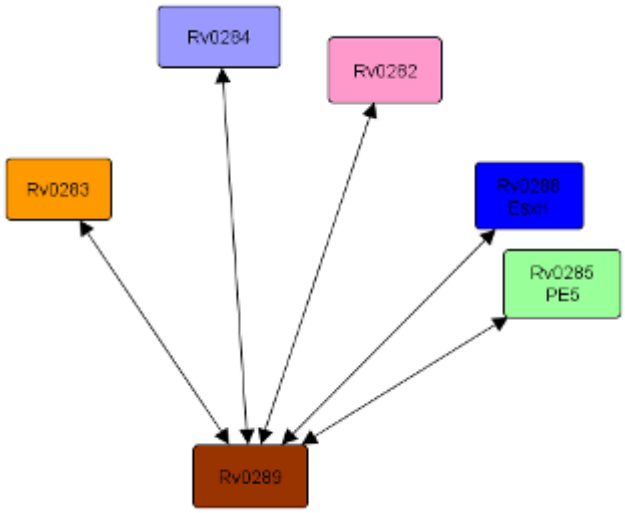


Figure 2.15. Rv0289 binds to several of the same proteins as the EsxG-EsxH complex

EsxH and PPE4 were also shown to interact. The interactions between PPE68 and ESAT-6, CFP-10 and EsxH have been shown previously (Okkels and Andersen, 2004; Teutschbein *et al.*, 2009). The putative ESX-3 substrates PE5-PPE4, EsxG-EsxH and Rv0289 interact with each other. The significance of these interactions is unclear, especially as they do not rely on each other for secretion. It has been proposed that rather than being substrates, Esx-Esx and PE-PPE form part of the ESX secretion machinery. It is also possible that the ESX substrates remain associated with the ESX secretion machinery, and the mycobacterial cell, via interactions with the mycomembrane-embedded PPE4.

The similarity and overlap of the interactions between EsxG and EsxH, and EsxQ, EsxR and EsxS, which were duplicated from EsxG and EsxH, suggest that these proteins may also be substrates of ESX-3, although this remains to be established. The regulation of expression of these proteins differs from that of EsxG and EsxH, in that it is not induced in response to iron starvation (Maciag *et al.*, 2007). Therefore EsxQ, EsxR and EsxS may be alternate ESX-3 substrates depending on the conditions in which the bacterium resides.

Limitations of this study

In this study we have identified several interactions which have not previously been identified for any of the other ESX secretion machineries, however some interactions previously identified between ESX-1 protein components have not been detected in the ESX-3 homologs.

The M-PFC technique has been specifically developed to investigate mycobacterial protein-protein interactions. Previous studies have utilised *in vitro* and yeast-2-hybrid techniques, which have lead to the exclusion of the identification of most interactions involving membrane proteins. In addition, for most of these approaches, the proteins are expressed in unrelated organisms, and may not undergo correct folding and post-translational modifications. Expression of mycobacterial proteins in *E. coli,* and other unrelated organisms, is notoriously difficult, often resulting in little or no expression, or the formation of inclusion bodies. This is mostly due to the high GC content of mycobacterial genes, as well as toxicity of some mycobacterial proteins. *M. smegmatis* is a heterologous mycobacterial host organism, which, being much more closely related to *M. tuberculosis*, has similar transcription and translation machinery, post-translational modification mechanisms and a similar membrane structure. Therefore, this technique is envisaged to be more efficient in identifying the protein-protein interactions of *M. tuberculosis,* specifically the membrane-associated interactions involved in secretion.

The technique is, however, not without limitations. There are several reasons why some protein-protein interactions may not have been identified during this study. These include the size of the proteins, the strength of the interaction, the conformation of the proteins, the location of the protein within the cell and competition from endogenous *M. smegmatis* proteins. The conformation of a protein, and its size, may affect the reconstitution of mDHFR, despite the presence of an interaction, thereby preventing the identification of the protein-protein interaction. Proteins which interact very weakly, or only interact transiently may produce very small amounts of mDHFR which is not sufficient to confer trimethoprim resistance to the bacteria. Protein conformation and interactions may also vary during different environmental conditions or growth phases. A protein pair may only interact under specific conditions and these interactions may remain undetected. In order for resistance to be conferred, both mDHFR fragments need to be located within the cell. As many of the ESX-3 components are membrane associated, one or

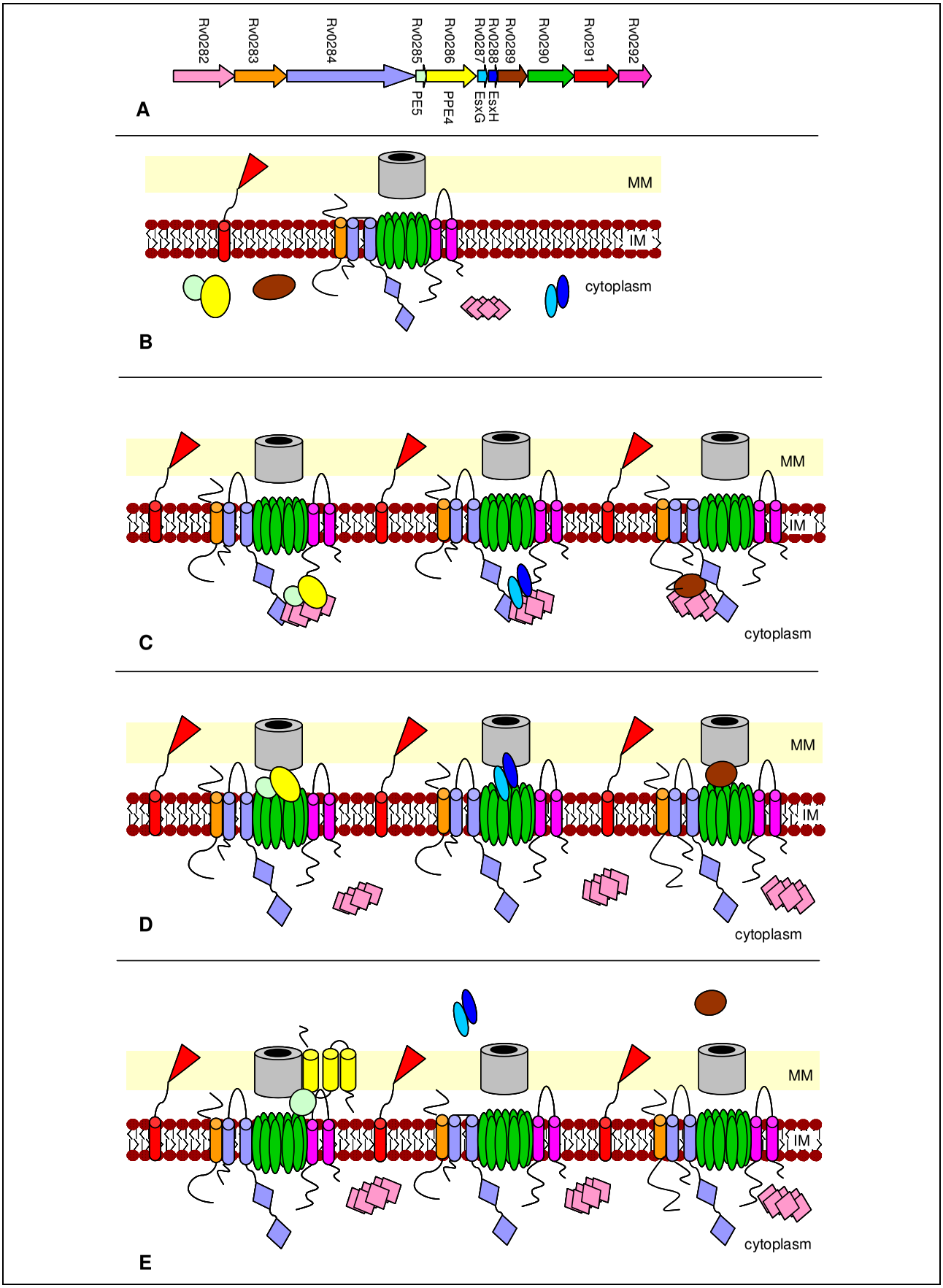
more of these fragments may be located outside of the cell, or within the membrane, preventing the detection of the interaction. As *M. smegmatis* has its own ESX-3 secretion system, some endogenous proteins may also interact with the mDHFR-fused proteins, preventing efficient interaction between the mDHFR fragments and preventing effective reconstitution of mDHFR, maintaining susceptibility to trimethoprim. In addition, some of the proteins may have N-terminal sequences which are cleaved post-translationally; should this occur, the N-terminal mDHFR fragment would also be cleaved and no interactions would be identified with that protein (mycosin-3 is thought to be autocatalyic, cleaving its N-terminal sequence, and this may explain why no interactions were detected for this protein (Rv0291)). Therefore some interactions which might occur between ESX-3 protein components may not have been identified within this study due to the above mentioned factors. Further selective cloning experiments of various sections of genes are needed to tease out the interactome to completion.

Non-specific interaction of proteins may also have lead to the identification of false interactions, therefore all interactions and their putative roles will need to be confirmed and further investigated. *In vivo* pull down assays, cross-linking studies and Blue Native PAGE (polyacrylamide gel electrophoresis) may be employed to further analyse the interactions of this secretion system.

General discussion

This is the first study which has attempted to identify protein-protein interactions and determine the interactome of the *M. tuberculosis* ESX-3 secretion system. Several ESX-3 protein-protein interactions have been identified. These interactions, together with the putative protein structures and localisations, and knowledge gained from other *M. tuberculosis* ESX secretion systems have been used to generate a model of ESX-3 secretion in *M. tuberculosis,* as proposed in Figure 2.16. This model and the interactions identified can serve as a basis for further investigation of the ESX-3 secretory mechanism. Confirming the interactions and roles of these proteins in ESX-3 secretion may help to elucidate the mechanism through which ESX-3 is involved in divalent metal cation homeostasis, an essential role in *M. tuberculosis*, and may provide novel drug targets for the eradication of tuberculosis disease.

Figure 2.16. A model of the ESX-3 secretion machinery. (**A**) The ESX-3 secretion machinery is encoded by the ESAT-6 gene cluster region 3. (**B**)The membrane proteins Rv0283, Rv0284, Rv0292 form a membrane complex with Rv0290, which forms the channel through the plasma membrane. The EsxG-EsxH, and PE5-PPE4 complexes form directly after translation. Rv0282 oligomerises to form a hexameric ATPase. (**C**) Rv0282 binds to EsxG-EsxH, PE5-PPE4 and Rv0289 and acts as a chaperone, providing energy to catalyse the interaction between these proteins and the membrane proteins Rv0284, and Rv0283 or Rv0292. (**D**) Rv0284 is an ATPase which provides the energy for translocation of the ESX-3 substrates through the Rv0290 membrane channel. (**E**) EsxG-EsxH and Rv0289 are secreted through the mycomembrane via an unknown pore protein, while PPE4 is embedded into the mycomembrane by its 3 hydrophobic transmembrane domains. PE5 may bind to Rv0290 to facilitate in associating PPE4 with the ESX-3 secretion machinery. The substrate(s) and role of mycosin 3 have yet to be determined, as does the identity of the mycomembrane channel protein. ATPase domains are represented as diamonds, IM is the inner/plasma membrane, MM is the mycomembrane.


Rv0282
Rv0283
Rv0284
Rv0285
Rv0286
Rv0287
Rv0288
Rv0289
Rv0290
Rv0291
Rv0292
PE5
PPE4
EsxH
EsxG
A
MM
IM
cytoplasm
B
MM
IM
cytoplasm
C
MM
IM
cytoplasm
D
MM
IM
cytoplasm
E

CHAPTER 3

The construction of a genetic knock-out of the ESX-3 secretion system in *Mycobacterium smegmatis*

3.1. Introduction

The *M. smegmatis* genome contains 3 of the 5 ESAT-6 gene cluster regions found in *M. tuberculosis*; regions 4, 3 and 1 (Gey van Pittius *et al*., 2001). The *M. smegmatis* ESAT-6 gene cluster region 3 contains all of the gene components of the *M. tuberculosis* gene cluster, spanning MSMEG_0615 to MSMEG_0626 (Figure 3.1) and encoding the ESX-3_{ms} secretion system. ESX-3 is essential for the *in vitro* growth of *M. tuberculosis* (Sassetti *et al*., 2003), however not for the growth of *M. smegmatis.*

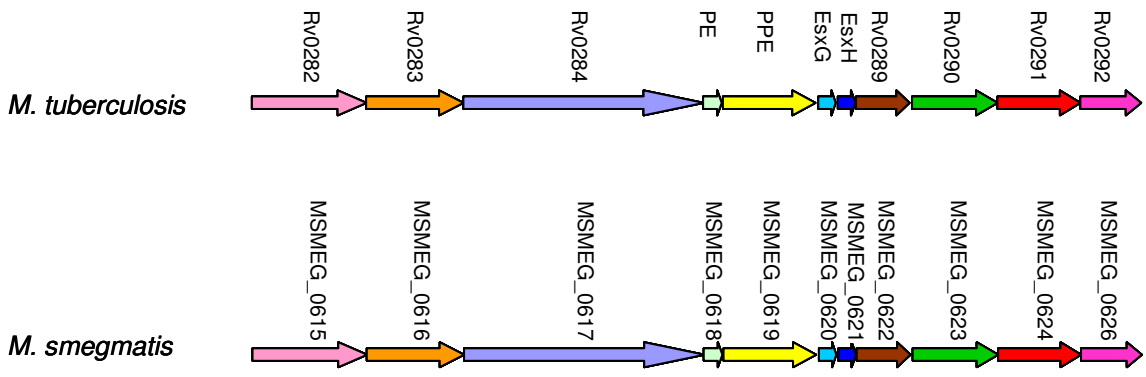


Figure 3.1. The ESAT-6 gene cluster region 3 of *M. tuberculosis* and *M. smegmatis*.

In this study we have made a knock-out of ESX-3_{ms} in *M. smegmatis*. This knock-out will enable the investigation of the functions, substrates and other properties of the *M. smegmatis* ESX-3 and allow further investigation into its essential function in *M. tuberculosis*. *M. smegmatis* is frequently used as a mycobacterial model organism to study *M. tuberculosis* due to its non-pathogenic nature, rapid growth rate and relative ease of manipulation. *M. smegmatis* and *M. tuberculosis* have a similar cytosolic composition, metabolism, mechanisms of gene expression and protein translation, folding, modifications and interactions. Therefore certain parallels may be drawn between what is observed in *M. smegmatis* and that which occurs in *M. tuberculosis* and other pathogenic mycobacteria. Understanding the function and mechanism of this secretion system may enable us to identify ways to interfere with it and to develop novel dugs to assist in eradicating tuberculosis disease.

3.2. Experimental approach

The ESAT-6 gene cluster region 3 (ESX-3_{ms}) was deleted from *M. smegmatis* using homologous recombination, as described by Gordhan and Parish (2001). Homologous recombination is a biological process during which two DNA strands with similar or identical sequence (homologous DNA) are broken and rejoined. This process is used in the repair of double stranded breaks during DNA replication and can be manipulated to incorporate, delete and exchange genetic material in the genomes of various organisms.

The technique used to delete the ESX-3_{ms} from *M. smegmatis* utilizes homologous recombination between a specifically constructed suicide vector and the genomic region of interest. The suicide vector is a non-replicating plasmid which contains an antibiotic selection marker as well as two DNA sequences identical to the DNA up- and downstream of the desired deletion region, located adjacent to one another. In addition, counter-selectable markers conferring sucrose-sensitivity and beta-galactosidase activity, may be included. A single recombination event between one of the homologous sequences and the genomic sequence results in the integration of the entire vector into the genome resulting in antibiotic resistant blue colonies (single cross-overs, SCOs) which are sensitive to sucrose. These SCOs are passaged to allow a second recombination event to occur between either of the homologous sequences resulting in double cross-overs (DCOs) which form sucrose resistant, antibiotic-sensitive white colonies. DCOs may result in either the removal of only the vector sequence (reversion to wild-type), or deletion of the desired region and the vector sequence (generating a knock-out); depending on the position of the second recombination event (Gordhan and Parish, 2001). Schematic outlines of the construction of the suicide vector and the homologous recombination events required to create a knockout of the *M. smegmatis* ESAT-6 gene cluster region 3 are given in Figure 3.2a and 3.2b, respectively.

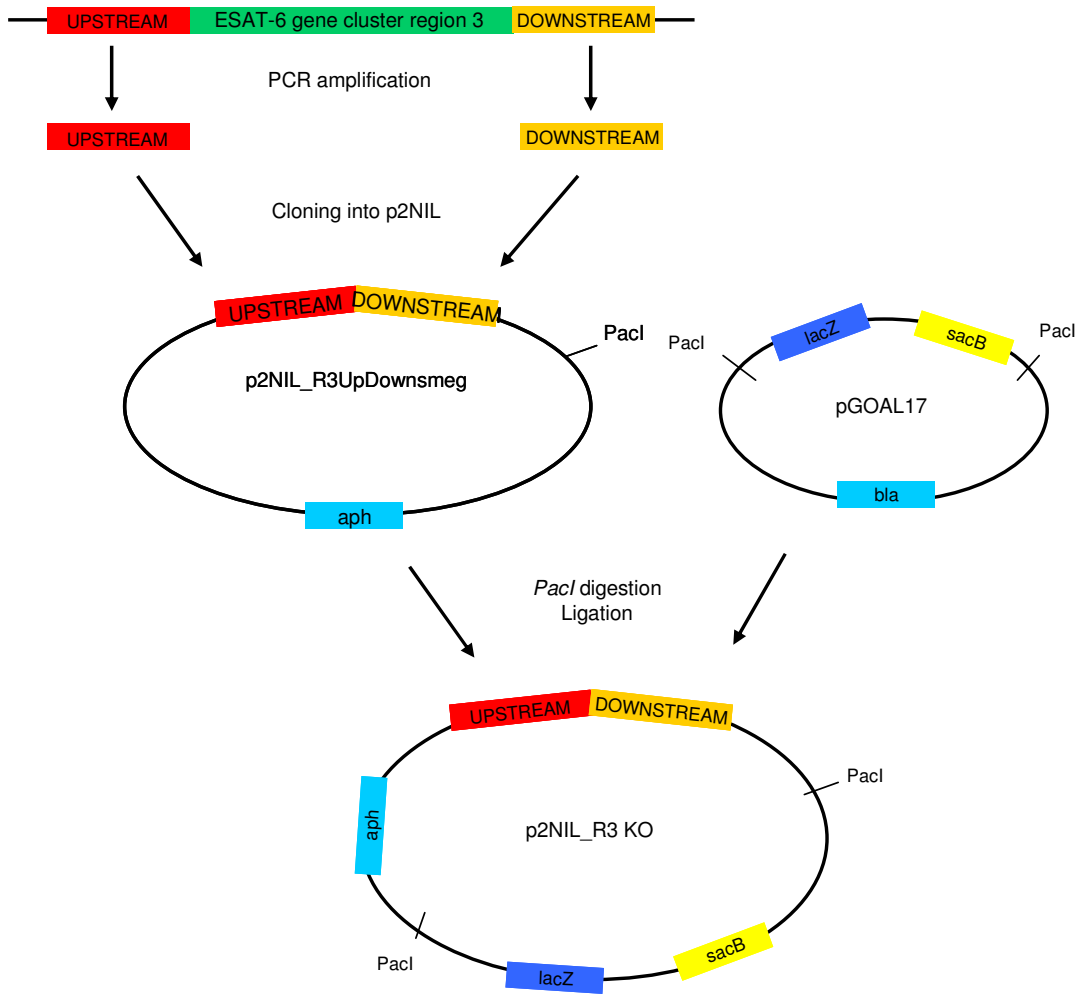


Figure 3.2a. Construction of the suicide vector p2NIL_R3 KO used to knock out the ESAT-6 gene cluster region 3 of *M. smegmatis*. DNA sequences of 800bp upstream and downstream of the ESAT-6 gene cluster region 3 are amplified from *M. smegmatis* genomic DNA and cloned into the suicide vector p2NIL to generate p2NIL_R3 UpDownsmeg. The *Pac*I cassette, containing the *sacB* and *lacZ* genes, is ligated into p2NIL_R3 UpDownsmeg to generate the knockout construct p2NIL_R3 KO. The *aph* and *bla* genes confer kanamycin and ampicillin resistance respectively.

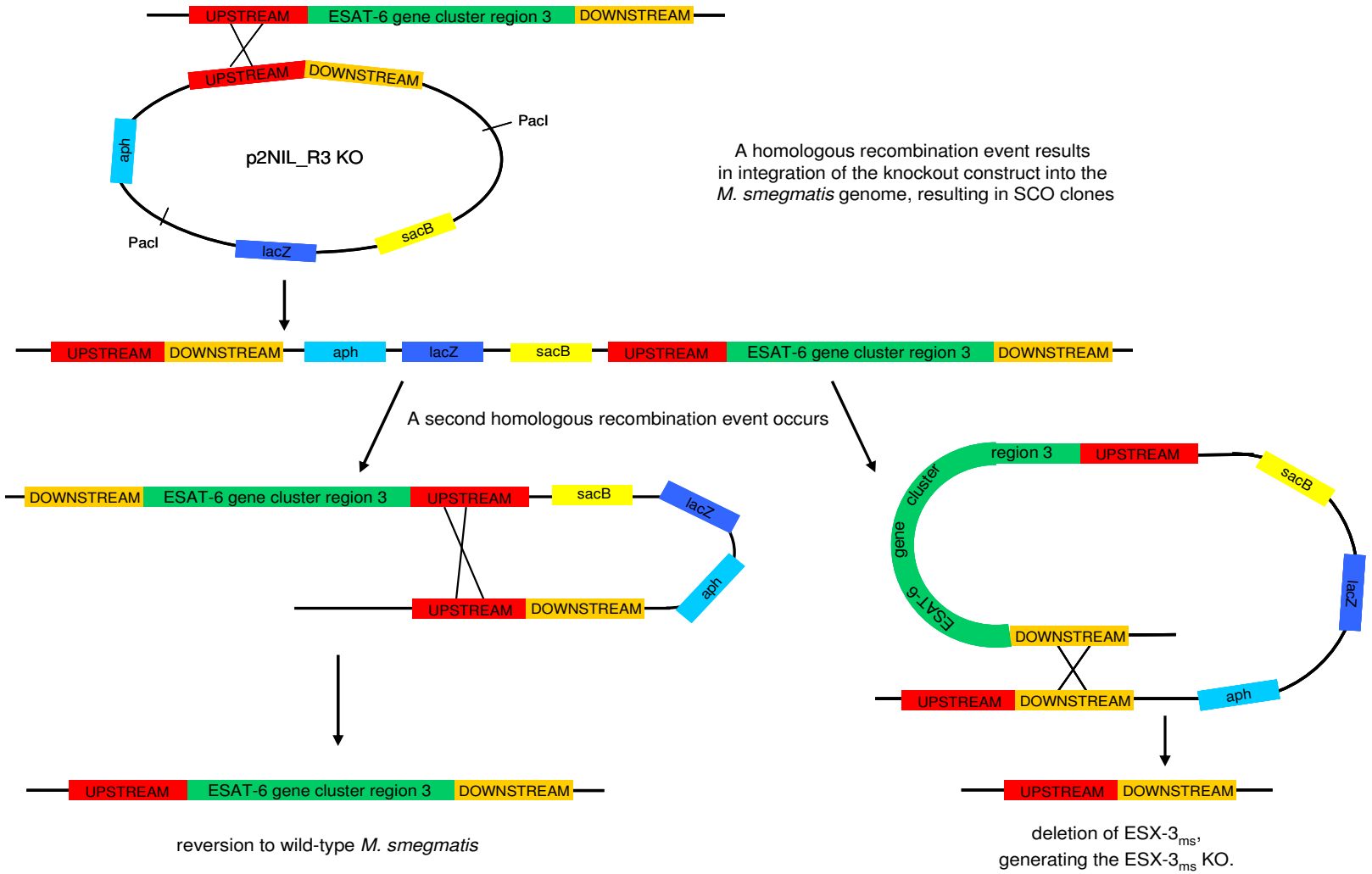


Figure 3.2b. Constructing a *M. smegmatis* ESAT-6 gene cluster region 3 knock-out by homologous recombination. The knock-out construct, p2NIL_R3 KO, is transformed into *M. smegmatis*. A single recombination event integrates the vector into the *M. smegmatis* genome, up or downstream of the ESAT-6 gene cluster region 3. A second recombination event is allowed to occur and results in either reversion to wild-type or generation of a knock-out of the ESAT-6 gene cluster region 3 ($ESX\text{-}3_{ms}$).

3.3. Materials and Methods

All standard molecular techniques were performed essentially as described by Sambrook *et al.,* (1989). Techniques and reagents are described in detail in Appendix A.

3.3.1. Bacterial strains

E. coli strain JM109 was used in cloning procedures. *M. smegmatis* mc^2155 was used to make the *M. smegmatis* ESAT-6 gene cluster region 3 knock-out (ESX-3_{ms} KO).

3.3.2. Media and culture conditions

E. coli was cultured in Luria-Bertani (LB) broth with shaking, and on LB agar plates, overnight at 37 ℃. Solid and liquid media were supplemented with the antibiotics ampicillin (50 ug/ml, Roche) and kanamycin (50 ug/ml, Sigma) and solid media with X-gal (Roche) and 5% sucrose, as appropriate.

M. smegmatis was grown in liquid LB with shaking, and on LB agar plates, for 4-5 days at 37 ℃. Liquid media was supplemented with 0.1% Tween-80, and kanamycin (25 ug/ml, Sigma) when required. Solid media was supplemented with kanamycin (25 ug/ml, Sigma), X-gal (Roche) and 5% sucrose, as appropriate.

3.3.3. Transformations

Transformation was performed by electroporation of electrocompetent cells. *E. coli* was transformed at 2.5 kV, 25 µF, 125 µFd, 200 Ω. One millilitre of SOC was added to the transformed cells which were expressed for 1 hour at 37 ℃ before plating. *M. smegmatis* was transformed at 2.5 kV, 25 µF, 125 µFd, 1000 Ω followed by the addition of 1 ml LB medium and expression for 3 hours at 37 ℃ before plating.

3.3.4. Cloning vectors

The desired PCR products were cloned into pGemT-Easy (Promega) and subcloned into the p2NIL mycobacterial suicide vector. The *Pac*I cassette containing the *sacB* and *lacZ* genes was digested from the pGOAL17 vector and cloned into the *Pac*I restriction site in p2NIL. The vectors used in this study are described in Table 3.1.

Table 3.1 Vectors used in the construction of the *M. smegmatis* ESX-3_{ms} KO.

Name	Description	Size (bp)	Source/Reference
pGEM-T easy	*E. coli* cloning T-vector, Amp^R, *lacZ*, *ori*E	3015	Promega
p2NIL	*E. coli* cloning vector, mycobacterial suicide vector, *ori*E, Kan^R	4753	Parish and Stoker 2000
pGOAL17	Plasmid with *lacZ* and *sacB* genes in PacI cassette, *ori*E, Amp^R	8855	Parish and Stoker 2000

Amp^R – ampicillin resistance; *lacZ* – β-galactosidase activity; *ori*E – *E. coli* origin of replication; Kan^R – kanamycin resistance; *sacB* – levansucrase activity

3.3.5. Primers

All *M. smegmatis* mc^2155 DNA sequence information was obtained from the JVCI CMR *Mycobacterium smegmatis* MC2 Genome Page (http://cmr.jcvi.org/tigr-scripts/CMR/GenomePage.cgi?org=gms). Primers R3 Upsmeg f & r and R3 Downsmeg f & r were designed to amplify approximately 800bp of DNA directly upstream and downstream of the ESAT-6 gene cluster region 3 of *M. smegmatis*, with *Kpn*I and *Bgl*II, and *Bgl*II and *Hind*III restriction sites respectively, to facilitate cloning into the suicide vector p2NIL. Additional primers, R3 F1 f & r and R3 KO f & r, were designed to distinguish between DCO colonies which may be wild-type *M. smegmatis* or the ESX-3_{ms} KO. The primers used in this study are described in Table 3.2 and Figure 3.3.

Table 3.2 Primers used in the construction of the *M. smegmatis* ESX-3_{ms} KO.

	Name	Sequence (5' – 3')	Length (bp)	Tm (°C)	Restriction site
A	R3 Upsmeg f	GGGGTACCGGAGCATCCGCTGCAGACC	27	64	*Kpn*I
B	R3 Upsmeg r	GGGGAGATCTCTCTCCCTTATGTATGCC	28	54	*Bgl*II
C	R3 Downsmeg f	GGGGAGATCTCGATCCCAGTGCTCCCACA	29	62	*Bgl*II
D	R3 Downsmeg r	GGGGAAGCTTCCCGAGCGATCCTTTCC	27	56	*Hind*III
E	R3 F1 f	GCAGTGGTTCTCCGAGCGTGG	21	70	n/a
F	R3 F1 r	ACGACGTCCGACCAGCGTTGG	21	70	n/a
G	R3 KO f	TCCTTCTTTGCGCTGGTCTT	20	60	n/a
H	R3 KO r	TGTCGCTGCCGTGGTTCT	18	58	n/a

Restriction enzyme recognition sites are underlined.

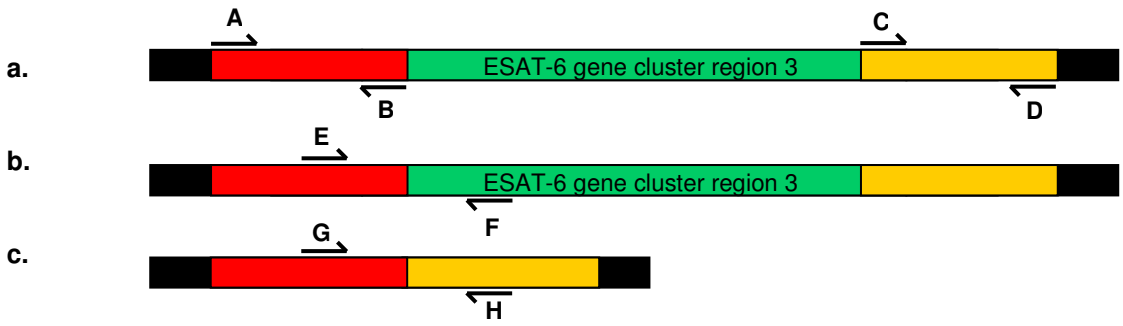


Figure 3.3. PCR primers used in the construction and verification of the *M. smegmatis* ESX-3_{ms} KO. (a.) Primers R3 Upsmeg f (A) and R3 Upsmeg r (B), and R3 Downsmeg f (C) and R3 Downsmeg r (D) were used to amplify 826bp of DNA sequence upstream and 829bp of DNA sequence downstream of the *M. smegmatis* ESAT-6 gene cluster region 3, respectively. (b.) The R3 F1 f (E) and R3 F1 r (F) primers amplify a 798bp DNA sequence upstream of and overlapping the ESAT-6 gene cluster region 3, and will generate a product in wild-type and SCO, but not ESX-3_{ms} KO *M. smegmatis*. (c.) R3 KO f (G) and R3 KO r (H) primers bind to sequences up and downstream of the ESAT-6 gene cluster region 3 and will amplify an 820bp DNA fragment only in *M. smegmatis* ESX-3_{ms} KO and SCO strains.

3.3.6. PCR amplification of R3 Upsmeg and R3 Downsmeg

M. smegmatis mc^2155 genomic DNA was used as a PCR template for the amplification of DNA sequences up- and downstream of the ESAT-6 gene cluster region 3 (ESX-3_{ms}). The DNA sequences upstream and downstream of the *M. smegmatis* gene cluster region 3 were PCR amplified with FastStart Taq DNA Polymerase (Roche), using the primers R3 Upsmeg f & r, and R3 Downsmeg f & r. The resultant products were named R3 Upsmeg and R3 Downsmeg respectively. The PCR products were separated by agarose gel electrophoresis, the appropriate bands excised and the DNA purified using the Wizard® SV PCR and Gel Cleanup Kit (Promega).

3.3.7. Cloning into pGemT-Easy

The purified R3 Upsmeg and R3 Downsmeg PCR products were ligated into the pGemT-Easy vector and transformed into *E. coli*. Transformants were selected for on LB agar plates with ampicillin (50 ug/ml) and PCR screened to verify the presence of the insert. Plasmids were extracted from 10 ml cultures using the Wizard® *Plus* SV Miniprep Plasmid Purification Kit, quantified and the inserts sequenced to confirm the sequence of the insert. The constructs were named pGemR3 Upsmeg and pGemR3 Downsmeg.

3.3.8. Construction of the *M. smegmatis* ESAT-6 gene cluster region 3 knockout construct.

pGemR3 Upsmeg and pGemR3 Downsmeg were digested with *Kpn*I and *Bgl*II, and *Bgl*II and *Hind*III respectively, simultaneously ligated into the *KpnI* and *HindIII* restriction sites of p2NIL, and transformed into *E. coli.* Transformants were selected on LB with kanamycin (50 ug/ml), and the construct was verified by colony PCR using the R3 Upsmeg f & r, R3 Downsmeg f & r and R3 Upsmeg f & R3 Downsmeg r primer sets and named p2NIL_R3 UpDownSmeg.

pGOAL17 was digested with *Pac*I, generating two bands of 2.3kb and 6.3kb each. The 6.3kb fragment, containing a *sacB* and *lacZ* gene was ligated into the *Pac*I site of p2NIL_R3 UpDownSmeg and transformed into *E. coli*. Blue colonies were selected on LB with kanamycin (50 ug/ml) and X-gal and verified by duplicate plating on LB with kanamycin (50 ug/ml) and 5% sucrose. The ESAT-6 gene cluster region 3 knockout construct, p2NIL_R3 KO, was isolated from clones which did not grow on sucrose-containing medium.

3.3.9. Generating the *M. smegmatis* ESAT-6 gene cluster region 3 knock-out

Wild-type *M. smegmatis* was transformed with 1 ug of the p2NIL_R3 KO plasmid and blue SCO colonies were selected on LB agar containing kanamycin (25 ug/ml) and X-gal. SCO colonies were picked and verified by lack of growth on LB containing 5% sucrose and colony PCR using primers R3 F1 f & r and R3 KO f & r. Both PCR reactions will amplify in SCOs.

Verified SCOs were passaged without selection in LB and plated on LB agar with 5% sucrose and X-gal and incubated at 37ºC for 3-4 days. White colonies, which are putative DCOs, were selected and inoculated into LB medium. Cultures were replica-plated on LB with 5% sucrose and X-gal, and LB with kanamycin (25 ug/ml) and X-gal to verify the DCO. DCOs will not grow on kanamycin-containing medium.

Colony PCRs were done on verified DCOs to distinguish between wild-type and ESX-3_{ms} KOs, using the primers R3 F1 f & r and R3 KO f & r. Wild-type *M. smegmatis* produces products for the PCR reaction R3 F1 but not for the R3 KO PCR. The ESX-3_{ms} KO will only produce a product for the R3 KO PCR. The R3 KO amplicon was sequenced to verify the knock-out.

3.4. Results

The 800bp DNA sequences upstream and downstream of the *M. smegmatis* ESAT-6 gene cluster region 3 were successfully PCR amplified from *M. smegmatis* mc^{2}155 genomic DNA (Figure 3.4a), cloned into pGemT-Easy and subcloned into the p2NIL suicide vector and confirmed by colony PCR (3.4b). This vector was called p2NIL_UpDownSmeg. Subsequent cloning of the pGOAL17 fragment containing the *lacZ* and *sacB* genes into p2NIL_R3 UpDownSmeg generated a construct, p2NIL_R3KO, which confers beta-galactosidase activity, kanamycin resistance and sucrose sensitivity to transformed cells.

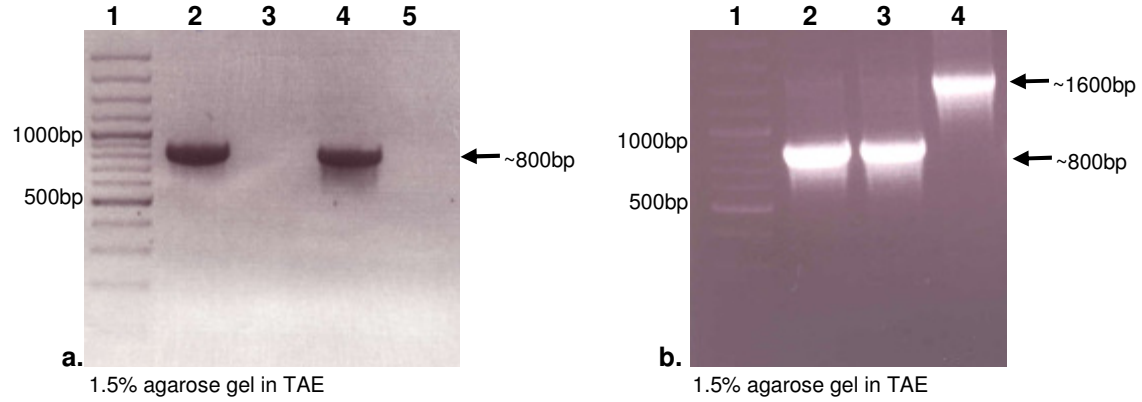


Figure 3.4. Construction of the *M. smegmatis* ESAT-6 gene cluster region 3 knock-out construct p2NIL_R3KO. **(a.)** The R3 Upsmeg (lane 2, 826bp) and R3 Downsmeg (lane 4, 829bp) PCR products were PCR amplified from *M. smegmatis* genomic DNA. Lane 1 is GeneRuler 100bp DNA Ladder Plus, lanes 3 and 5 are negative controls. (**b.)** The p2NIL_R3 UpDownsmeg construct contains both R3 Upsmeg (lane 2, 826bp) and R3 Downsmeg (lane 3, 829bp), situated adjacent to one another (lane 4, 1655bp).

The p2NIL_R3 KO construct was transformed into *M. smegmatis* mc^{2}155 resulting in blue SCO colonies on LB medium containing X-gal and kanamycin. LB medium was used for the growth of all cultures used to generate the knock-out, as Middlebrook 7H9 and 7H11 media contain dextrose which is converted by the product of the *sacB* gene, levansucrase, into a toxic product, preventing the growth of the bacteria. The SCO colonies were confirmed by PCR (Figure 3.5) as well as replica plating on LB medium with and without sucrose to confirm the presence of the *sacB* gene. SCOs did not grow on the plates containing sucrose.

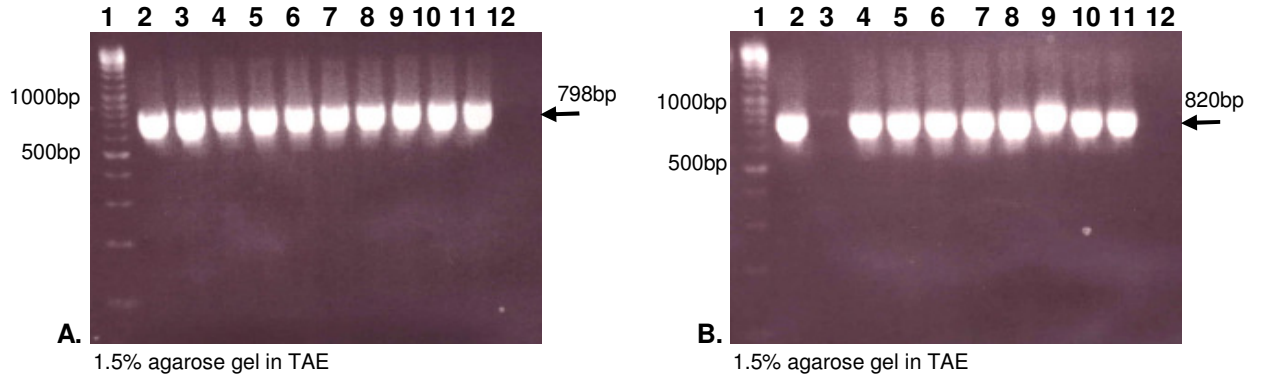


Figure 3.5. *M. smegmatis* SCO colonies generated during the construction of the *M. smegmatis* ESX-3$_{ms}$ KO. Colony PCR was used to verify putative SCO colonies. SCOs are positive for amplification of 798bp PCR products with primers R3 F1 f & r (A) and 820bp PCR products with primers R3 KO f & r (B). Lanes 2, 4, 5, 6, 7, 8, 9, 10 and 11 contain SCOs. Lane 1 contains GeneRuler 100bp DNA Ladder Plus and Lane 12 is a negative control.

SCOs were passaged without selection to allow a second recombination event to occur. Cultures were plated on LB medium containing sucrose and X-gal to select for clones without the *aph, lacZ* and *sacB* genes which are lost during a second cross-over event. White colonies were selected and replica plated on LB with sucrose and X-gal and LB with kanamycin. Clones which formed white colonies on sucrose-containing medium, and showed no growth on kanamycin were confirmed as DCOs. PCR was used to distinguish between wild-type and ESAT-6 gene cluster region 3 knock-out (ESX-3$_{ms}$ KO) DCOs. Two of the isolated DCOs were found to be ESX-3$_{ms}$ KOs (Figure 3.6). Sequencing confirmed the absence of the ESAT-6 gene cluster region 3 from these cultures. Glycerol freezer stocks of these cultures were made and stored at -80ºC for future applications.

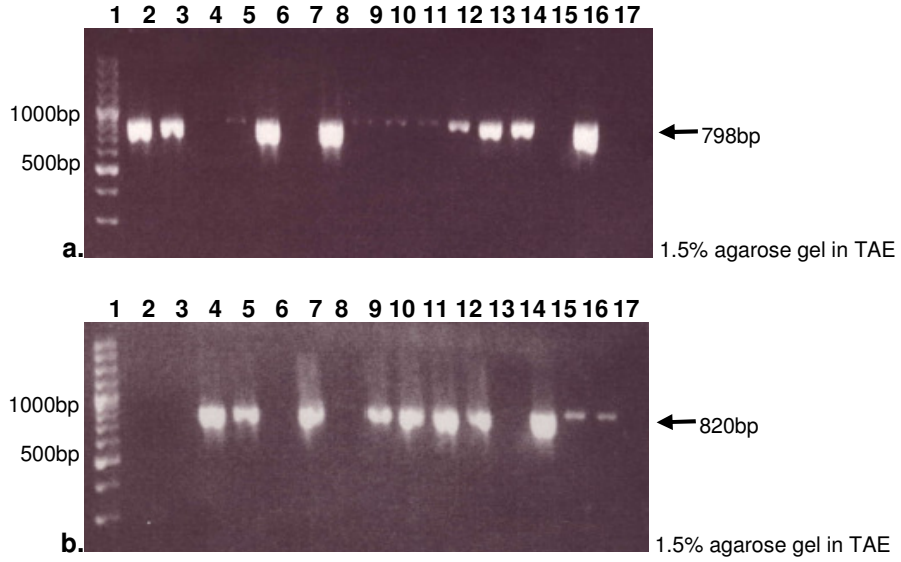


Figure 3.6. Identification of the *M. smegmatis* ESX-3_{ms} KO. Colony PCR was used to distinguish between DCO colonies which contained wild-type and R3 KO *M. smegmatis.* Wild-type *M. smegmatis* produces a 798bp PCR product with primers R3 F1 f & R **(a).** R3 KOs produce 820bp PCR products with primers R3 KO f & r **(b).** Lane 1 contains GeneRuler 100bp DNA Ladder Plus. Lanes 2, 3, 6, 8 and 13 are wild-type *M. smegmatis*. Lanes 4 and 7 are *M. smegmatis* R3 KOs. Lane 2 is an *M. smegmatis* wild-type control. Lane 17 is a negative control. The remaining lanes are SCOs, mixed wild-type and R3 KO colonies or inconclusive PCR results.

3.5. Discussion

This study describes the construction of an *M. smegmatis* ESAT-6 gene cluster region 3 knock-out strain, *M. smegmatis* ESX-3_{ms} KO, using homologous recombination. The construction of this knock-out strain confirms the non-essential nature of the ESAT-6 gene cluster region 3 in *M. smegmatis*. This knock-out strain may be used in various comparative studies with the *M. smegmatis* wild-type strain to identify differences in their proteomes, metabolomes and secretomes, as well as other properties, thereby enabling the identification of functions and substrates of the *M. smegmatis* ESX-3 secretion system, which could be informative regarding the *M. tuberculosis* ESX-3. It also provides an opportunity to study inter-region interactions. This may highlight reasons for the essential nature of this secretion system in *M. tuberculosis* and allow us to expand our knowledge and understanding of the immunopathologically important ESX secretion systems and the mechanisms of pathogenesis and survival of *M. tuberculosis,* and may help to identify new drug and vaccine targets for the treatment and prevention of tuberculosis.

Recently, a conditional ESX-3 mutant of *M. tuberculosis* was constructed (Serafini *et al.*, 2009). Repression of ESX-3 expression in this mutant was lethal to the bacterium, however the phenotype can be complemented by zinc, iron and wild-type culture supernatant. This clearly indicates that *M. tuberculosis* ESX-3 secretes factors which are involved in iron and zinc uptake. *M. tuberculosis* ESX-3 expression is regulated by iron and zinc (Rodriguez *et al.*, 2002; Maciag *et al.*, 2007), supporting this role for ESX-3 in metal cation homeostasis. The fact that *M. smegmatis* ESX-3 expression is also regulated by iron suggests a level of conservation of function between species (Maciag *et al.*, 2009). Therefore the *M. smegmatis* ESX-3_{ms} knock-out may enable further investigation of ESX-3's role in metal cation homeostasis.

Chapter 4 investigates the secretion of *M. tuberculosis* EsxH by the *M. smegmatis* ESX-3_{ms} KO strain in comparison to the *M. smegmatis* wild-type strain in an attempt to characterize the role of ESX-3_{ms} in EsxH secretion, and the functional conservation of the ESX-3 secretion system between the two mycobacterial species. In addition, the *M. smegmatis* ESX-3_{ms} KO is currently being used in a comparative study with the *M. smegmatis* wild-type strain to identify variations in their metabolomes, in collaboration with Dr Loots at North-West University. This will form the basis of further studies.

CHAPTER 4

Investigation of *Mycobacterium tuberculosis* EsxH secretion by the *Mycobacterium smegmatis* ESX-3 secretion system

4.1. Introduction

The *M. smegmatis* genome contains three of the ESAT-6 gene clusters present in the *M. tuberculosis* genome, namely ESAT-6 gene cluster regions 4, 3 and 1(Gey van Pittius *et al*., 2001), each encoding an ESX secretion system. The *M. tuberculosis* ESAT-6 gene clusters encode several immunogenic Esx proteins which have been identified in the culture filtrates of *M. tuberculosis*; including EsxA (ESAT-6) and EsxB (CFP-10) from region 1, and EsxH (TB10.4) from region 3 (Andersen *et al*., 1995; Berthet *et al*., 1998; Skjot *et al*., 2000).

The ESX-1 secretion system is responsible for secretion of the EsxAB (ESAT-6/CFP-10) protein complex. Converse and Cox (2004) have shown that the *M. smegmatis* ESX-1 secretion system is able to secrete *M. tuberculosis* EsxAB. This secretion system is not involved in virulence in the saprophytic *M. smegmatis,* but has rather been shown to be involved in conjugal DNA transfer (Flint *et al*., 2004; Coros *et al*., 2008), which has not been observed in *M. tuberculosis*. Despite this, the *M. tuberculosis* ESX-1 can complement the ESX-1 DNA conjugation phenotype in an *M. smegmatis* ESX-1 mutant (Flint *et al*., 2004). These data indicate that the ESX-1 secretion system is functionally conserved between these two species and may perform additional, as yet unidentified functions in both species.

Both *M. smegmatis* and *M. tuberculosis* also contain the ESAT-6 gene cluster region 3, encoding ESX-3. Encoded in this region are the Esx family proteins EsxG and EsxH, which form an EsxGH complex (Okkels and Andersen, 2004). EsxH was identified in the culture filtrates of *M. tuberculosis* and initially named TB10.4 (Skjot *et al*., 2000). EsxH is a potent immunogen, suggesting that it is highly expressed during infection. To determine whether the ESX-3 secretion systems of *M. tuberculosis* and *M. smegmatis* are functionally conserved, we investigated the secretion of *M. tuberculosis* EsxGH by the *M. smegmatis* wild type and ESX-3_{ms} knock-out strains, and those complemented with the *M. tuberculosis* ESX-3.

ESX-3 is essential in *M. tuberculosis* (Sassetti *et al*., 2003), but not in *M. smegmatis*, suggesting that this secretion system either performs different functions in the two species, or that the two species have different requirements for survival. *M. tuberculosis* ESX-3 expression is regulated by iron and zinc (Rodriguez *et al*., 2002; Maciag *et al*., 2007), while *M. smegmatis* ESX-3 is only regulated by iron (Maciag *et al*., 2009), indicating that at least regulation of ESX secretion in these two species is not identical.

Determining whether *M. smegmatis* ESX-3 is able to secrete *M. tuberculosis* EsxGH may be a first step toward establishing whether ESX-3 is also functionally conserved between these two species, which will help to validate the use of the *M. smegmatis* ESX-3 knock-out to determine the function of ESX-3 in *M. tuberculosis.*

4.2. Experimental approach

This study investigates the secretion of *M. tuberculosis* EsxH (TB10.4) by the *M. smegmatis* ESX-3 secretion system. Western blotting, employing a mouse anti-TB10.4 antibody is used to identify EsxH in the whole cell lysate (WCL) and culture filtrate (CF) fractions of *M. smegmatis* wild-type and ESX-3_{ms} KO strains expressing either EsxG and EsxH, or the entire *M. tuberculosis* ESX-3 secretion system.

4.3. Materials and Methods

All standard molecular techniques were performed essentially as described by Sambrook *et al.,* (1989). Techniques and reagents are described in detail in Appendix A.

4.3.1. Bacterial strains

M. smegmatis mc^2155 (Snapper *et al*., 1990) and the *M. smegmatis* mc^2155 ESX-3_{ms} knock-out strain (Chapter 3) were used for EsxH secretion analyses.

4.3.2. Media and culture conditions

M. smegmatis was grown in Middlebrook 7H9 medium with shaking, and on BBL™ Seven H11 Agar Base plates, at 37 ℃ for 3-4 days or as otherwise stated. Solid and liquid media were supplemented with 0.5% glycerol, 0.5% glucose and 0.2% Tween-80. Hygromycin (50 ug/ml) was added as appropriate.

For secretion studies cultures were grown in Kirchener's minimal medium supplemented with hygromycin (50 ug/ml) when necessary.

4.3.3. Construction of *M. smegmatis* strains for the TB10.4 secretion study

The pYUB412_ESX-3_{mtb}, pUAB300_Rv0287-0288 and p19Kpro_EsxGH vectors (Table 4.1) were individually transformed into *M. smegmatis* and *M. smegmatis* ESX-3_{ms} KO and plated on 7H11 plates containing hygromycin. Colonies were picked and inoculated into 7H9 medium with hygromycin and incubated at 37 ℃ with shaking for 2 days. *M. smegmatis* and *M. smegmatis* ESX-3_{ms} KO were grown in 7H9 without hygromycin. The *M. smegmatis* WT and ESX-3_{ms} KO strains, and those containing pUAB300_Rv0287-0288, p19Kpro_EsxGH and pYUB412_ESX-3_{mtb} were confirmed by PCR using the R3 KO f & r, R3 F1 f & r and F102 & R102, p19Kpro_EsxG sense & p19Kpro_EsxH antisense, and Rv0282 f & r, Rv0285 f & r, Rv0292 f & r primers, respectively (Table 4.2). The different strains constructed and used in this study are described in Table 4.3.

Table 4.1. Vectors used in the EsxH secretion study

Name	Description	Size (bp)	Source/ Reference
pYUB412_ESX-3_{mtb}	Integrative cosmid containing 181kb *M. tuberculosis* DNA including the ESAT-6 gene cluster region 3. *ori*E, cosA, cosB, attP, int, Hyg^R, Amp^R	189744	A kind gift from N.C. Gey van Pittius
pUAB300_Rv0287-0288	*E. coli*-Mycobacterial shuttle vector containing Rv0287 and Rv0288, episomal, Hyg^R, groEL promoter, DHFR (1,2), *ori*M, *ori*E	5603	Chapter 2
p19Kpro_EsxGH	*E. coli*-Mycobacterial shuttle expression vector containing EsxG and EsxH (Rv0287 and Rv0288) Hyg^R, *ori*E, *ori*M, 19 kDa antigen promoter	6153	A kind gift from Z. Fang

*ori*E – *E. coli* origin or replication; cosA – cohesive end site A; cosB – cohesive end site B; ; attP – attachment site of phage; int – integrase; Hyg^R – hygromycin resistance; Amp^R – ampicillin resistance; DHFR – dihydrofolate reductase fragment; *ori*M – Mycobacterial origin of replication

Table 4.2. Primers used in the EsxH secretion study.

Name	Sequence (5' – 3')	Length (bp)	Tm (°C)	Product size (bp)
R3 F1 f	GCAGTGGTTCTCCGAGCGTGG	21	70	798 (WT only)
R3 F1 r	ACGACGTCCGACCAGCGTTGG	21	70	
R3 KO f	TCCTTCTTTGCGCTGGTCTT	20	60	820 (ESX-3_{ms} KO only)
R3 KO r	TGTCGCTGCCGTGGTTCT	18	58	
F102	AGAACCACCACGAGGAGCTCAT	22	60.3	800
R102	TGATGCCTGGCAGTCGATCGTA	22	60.4	
Rv0282 pUAB300 f	ATCGATGATGGCGGGCGTAGGTGAA	25	64.0	1930
Rv0282 pUAB300/400 f	ATCGATCGTGGTCGTGCTGCTGGTT	25	64.4	
Rv0285 pUAB300 f	ATCGATCATGACGTTGCGAGTGGTTCC	27	62.4	325
Rv0285 pUAB300/400 f	ATCGATTGCTCAGCCGCCCACGAC	24	65.6	
Rv0292 pUAB300 f	ATCGATGGGAGCCCACCGAATGAAC	25	62.9	1059
Rv0292 pUAB300/400 f	ATCGATGGACTATCTGCGGCGGATGA	26	63.2	
esxGp19 sense	GGATCCAGGCGGAGATGTTATGAGCCT TTTGGA	33	65.3	654
esxHp19 antisense	ATCGATTTAGTGGTGGTGGTGGTGG GCCCCATTTGGCGGCTTCGG	47	73.5	

Table 4.3. *M. smegmatis* strains constructed and used in the EsxH secretion study

Name	Description	Source/Reference
M. smegmatis WT	Wildtype *M. smegmatis* mc^2155	(Snapper *et al*., 1990)
M. smegmatis WT + ESX-3_{mtb}	Wildtype *M. smegmatis* mc^2155 containing pYUB412_ESX-3_{mtb}	This study
M. smegmatis WT + EsxGH	Wildtype *M. smegmatis* mc^2155 containing pUAB300_Rv0287-88	This study
M. smegmatis ESX-3_{ms} KO	*M. smegmatis* mc^2155 ESX-3_{ms} KO	Chapter 3
M. smegmatis ESX-3_{ms} KO + ESX-3_{mtb}	*M. smegmatis* mc^2155 ESX-3_{ms} KO containing pYUB412_ESX-3_{mtb}	This study
M. smegmatis ESX-3_{ms} KO + EsxGH	*M. smegmatis* mc^2155 ESX-3_{ms} KO containing pUAB300_Rv0287-88	This study
M.smegmatis WT + EsxGH-His	*M. smegmatis* mc^2155 WT containing p19Kpro_EsxGH	This study
M.smegmatis ESX-3_{ms} KO + EsxGH-His	*M. smegmatis* mc^2155 ESX-3_{ms} KO containing p19Kpro_EsxGH	This study

4.3.4. Culturing of *M. smegmatis* for secretion analyses

Each of the *M. smegmatis* strains used for the secretion analysis (Table 4.3) was inoculated into 10 ml 7H9 with hygromycin, when appropriate. Cultures were grown, with shaking, at 37 ℃ to mid-log phase (OD_{600} = 0.8). Cultures were washed twice with Kirchener's medium with 0.2% Tween-80, and inoculated into 100 ml fresh Kirchener's medium without Tween-80, to OD_{600} = ~0.05. Cultures were grown, with shaking, at 37 ℃ to OD_{600} = ~0.4, to prevent the lysis of cells in culture.

4.3.5. Secretion analysis

M. smegmatis cultures were fractionated into the cell pellet and culture supernatant fractions by centrifugation. The whole cell lysate was extracted by mechanical cell lysis, by ribolyzing, and the proteins from the culture filtrate were acetone precipitated, as described in Addendum A. His-tagged proteins were purified from cultures containing p19Kpro_EsxGH using the MagneHis™ Protein Purification System (Promega) essentially according to manufacturers instructions, as described in Addendum A.

The concentrations of the protein fractions were determined spectrophotometrically using the Bio-Rad Protein Assay as described by the manufacturer, according to the microtitre plate protocol. All samples were read in duplicate, at three dilutions and the protein concentrations determined using a Bovine Serum Albumin (BSA) protein standard curve.

The culture filtrate and whole cell lysate proteins were separated by SDS-PAGE (sodium dodecyl sulphate – polyacrylamide gel electrophoresis). Five micrograms of each whole cell lysate sample, and 2 ug of each culture filtrate sample, or 10 ul of His-purified protein were loaded on a 3-18% SDS-PAGE gel and electrophoresed at 150 V. Purified recombinant TB10.4 protein (1 ug, A kind gift from P. Anderson) was used as a positive control and purified recombinant GroEL protein (1 ug, Colorado State University (CSU)) as a negative control and to eliminate false positive results due to cell lysis. Gels were run in duplicate. One gel was stained with silver nitrate to visualise the proteins. Western blotting was done on the second gel, using mouse anti-TB10.4 primary antibody (1 in 12 dilution, A kind gift from P. Anderson) or anti-His (C-terminal) antibody (1 in 5000 dilution, Invitrogen), and goat anti-mouse IgG-HRP (horseradish peroxidase) secondary antibody (1 in 5000 dilution, Invitrogen) and detected using the ECL Plus Western Blotting Detection System (Amersham™). Membranes were stripped and reprobed using the HisDetector™ Western Blot Kit, HRP Colorimetric detection kit (Kirkegaard & Perry Laboratories (KPL)), as described by the manufacturer, when appropriate.

4.4. Results

4.4.1. Construction of *M. smegmatis* strains for EsxH secretion analysis

The pYUB412_ESX-3_{mtb}, pUAB300_Rv0287-0288 and p19Kpro_EsxGH constructs were transformed into the *M. smegmatis* WT and ESX-3_{ms} KO strains to create the following strains; *M. smegmatis* WT + ESX-3_{mtb}, *M. smegmatis* WT + EsxGH, *M. smegmatis* ESX3 KO + ESX-3_{mtb}, *M. smegmatis* ESX-3_{ms} KO + EsxGH. Transformants were selected on 7H11 medium with hygromycin and PCR screened to confirm the relevant genotypes (Figure 4.1).

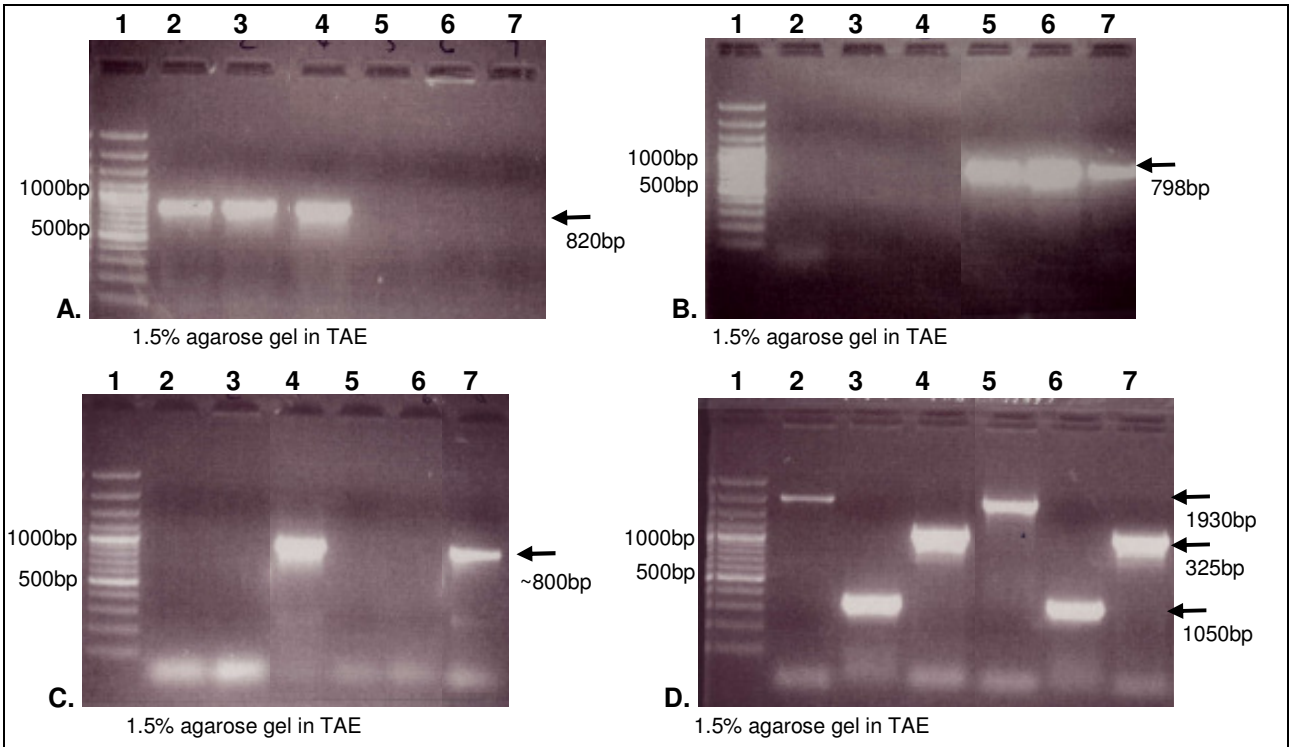


Figure 4.1. PCR confirmation of the *M. smegmatis* strains used for EsxH secretion analysis. The *M. smegmatis* strains were confirmed by PCR using various primer sets; **A**. R3 KO f&r, **B**. R3 F1 f&r, **C**. F102 & R102. Lane 1 is GeneRuler 100bp DNA Ladder Plus (Fermentas); Lane 2-4 are *M. smegmatis* ESX-3_{ms} KO containing either no vector (2), pYUB412_ESX-3_{mtb} (3) or pUAB300_Rv0287-0288 (4); Lane 5-7 are *M. smegmatis* WT containing either no vector (5), pYUB412_ESX-3_{mtb} (6) or pUAB300_Rv0287-0288 (7). **D**. The presence of the *M. tuberculosis* ESX-3 in *M. smegmatis* strains containing pYUB412_ESX-3_{mtb} confirmed by PCR amplification of Rv0282, Rv0285 and Rv0292 using primer sets Rv0282 pUAB300 f & Rv0282 pUAB300/400 r (Lane 2 and 5, ~1930bp), Rv0285 pUAB300 f & Rv0285 pUAB300/400 r (Lane 3 and 6, ~325bp) and Rv0292 pUAB300 f & Rv0292 pUAB300/400 r (Lane 4 and 7, ~1050bp). Lane 1 is GeneRuler DNA Ladder Plus (Fermentas); Lane 2-4 *M. smegmatis* ESX-3_{ms} KO + ESX-3_{mtb}; Lane 5-7 *M. smegmatis* WT + ESX-3_{mtb}

4.4.2. Secretion analysis of EsxH expressed from pUAB300 and pYUB412

We investigated the expression of EsxH from the mycobacterial M-PFC expression vector pUAB300 and the pYUB412 cosmid containing the *M. tuberculosis* ESAT-6 gene cluster region 3. The whole cell lysates of the *M. smegmatis* WT and ESX-3_{ms} KO cultures expressing ESX-3_{mtb} and EsxH were separated by SDS-PAGE. The gels were stained with silver stain (Figure 4.2A) or analysed by Western blotting using a mouse anti-TB10.4 primary antibody and an HRP-conjugated goat anti-mouse secondary antibody to determine whether EsxH protein is expressed in these cultures (Figure 4.2B). Silver staining of the SDS-PAGE gel showed that all of the WCLs contained a range of protein band sizes, and the TB10.4 protein control was present as two bands of just under 15 kDa. All six WCL samples contained protein bands of approximately this size. Western blotting of the duplicate gel strongly detected the TB10.4 positive control protein, however no EsxH protein was detected in any of the samples.

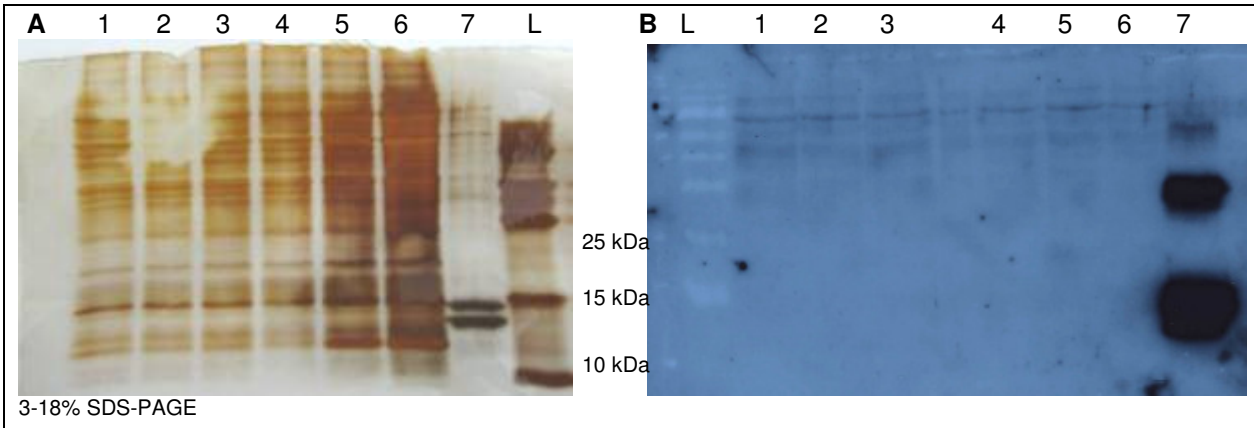


Figure 4.2. EsxH was not detected in the *M. smegmatis* WCLs. A. SDS-PAGE of the WCL fractions of the *M. smegmatis* strains followed by Silver staining (A) and Western Blotting (B) showed that although the WCLs were rich in proteins of a range of sizes, no TB10.4 protein could be detected in any of the samples. L. PageRuler Prestained Protein Ladder (Fermentas), 1. *M. smegmatis* WT, 2. *M. smegmatis* WT + ESX-3_{mtb}, 3. *M. smegmatis* WT + EsxGH, 4. *M. smegmatis* ESX-3_{ms} KO, 5. *M. smegmatis* ESX-3_{mtb} KO + ESX-3_{mtb}, 6. *M. smegmatis* ESX-3_{ms} KO + EsxGH, 7. TB10.4 positive control.

No EsxH could be detected in the WCLs of any of the *M. smegmatis* wild-type or ESX-3_{ms} knock-out strains with or without the *esxG-esxH* operon, although the recombinant TB10.4 positive control was detected by Western blotting. This suggests that there is no crossreactivity between the *M. smegmatis* Esx proteins and the TB10.4 primary antibody. It appears that expression of EsxH from the ESAT-6 gene

cluster region 3 contained in pYUB412_ESX3 as well as from pUAB300 was too low to be detected by Western Blotting.

4.4.3. Construction of *M. smegmatis* strains for His-tagged EsxH secretion analysis

In an attempt to express larger amounts of EsxH in *M. smegmatis*, and enable more sensitive detection thereof, in order to establish whether *M. smegmatis* ESX-3_{ms} can secrete *M. tuberculosis* EsxH, we coexpressed EsxG and EsxH from the mycobacterial expression vector p19Kpro_EsxGH. The EsxH protein expressed from this vector contains a C-terminal Histidine tag to enable the protein to be enriched for from the WCL and CF. This vector was transformed into *M. smegmatis* WT and *M. smegmatis* ESX-3_{ms} KO strains to produce *M. smegmatis* WT + EsxGH-His and *M. smegmatis* ESX-3_{ms} KO + EsxGH-His, to determine whether *M. smegmatis* ESX-3_{ms} is able to secrete *M. tuberculosis* EsxH. The genotypes of the transformants were confirmed by PCR (Figure 4.3).

4.4.4. Secretion analysis of His-tagged EsxH from *M. smegmatis* WT and ESX-3_{ms} KO strains

M. smegmatis WT + EsxGH-His and *M. smegmatis* ESX-3_{ms} KO + EsxGH-His strains were cultured and the His-tagged EsxH protein was enriched for from the WCLs using the MagneHis™ Protein Purification System. The purified EsxH-His proteins were separated by SDS-PAGE. Silver staining detected a band in each of the WCL fractions, corresponding to the size of one of the TB10.4 positive control bands (Figure 4.4A). Western blotting of a duplicate gel, using mouse anti-TB10.4 primary antibody and goat anti-mouse IgG-HRP secondary antibody detected only the TB10.4 positive control (Figure 4.4B). No His-tagged protein could be detected in the WCL fractions by reprobing with anti-His (C-terminal) primary antibody and goat anti-mouse IgG-HRP (horseradish peroxidase) secondary antibody, or the HisDetector™ HRP Colorimetric Western Blot Detection Kit (not shown). We predict that the level of expression of EsxH from p19Kpro is insufficient to allow detection of this protein by any of the Western blotting methods used and that the proteins observed by SDS-PAGE and silver staining after His purification are background. These results are supported by independent analyses by Mr Z. Fang looking at the expression of various His-tagged *M. tuberculosis* proteins, including EsxG and EsxH from p19Kpro, where no His-tagged protein could be specifically detected by Western Blotting.

As no EsxH, or His-tagged EsxH protein could be detected in any of the WCLs, no further analyses could be done.

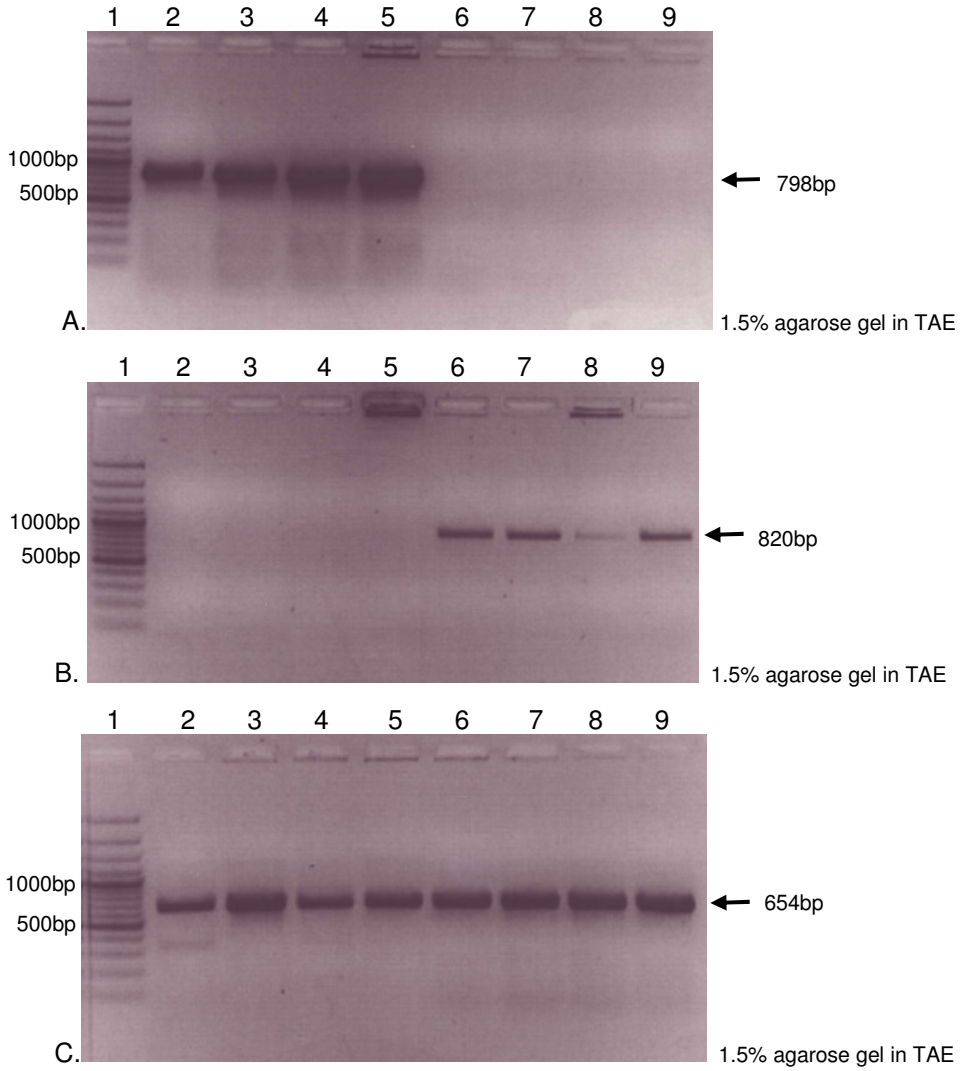


Figure 4.3. PCR confirmation of the *M. smegmatis* WT and ESX-3_{ms} KO strains containing p19Kpro_EsxGH. Colonies were screened to confirm *M. smegmatis* WT + EsxGH-His (Lane 2-5) and *M. smegmatis* ESX-3_{ms} KO + EsxGH-His (Lane 6-9). PCR was performed using the primersets (A.) R3 F1 f&r (798bp in WT), (B.) R3 KO f&r (820bp in ESX-3_{ms} KO) and (C.) p19Kpro_EsxG sense & p19Kpro_EsxH antisense (654bp). Lane 1 contains GeneRuler 100bp DNA Ladder Plus (Fermentas).

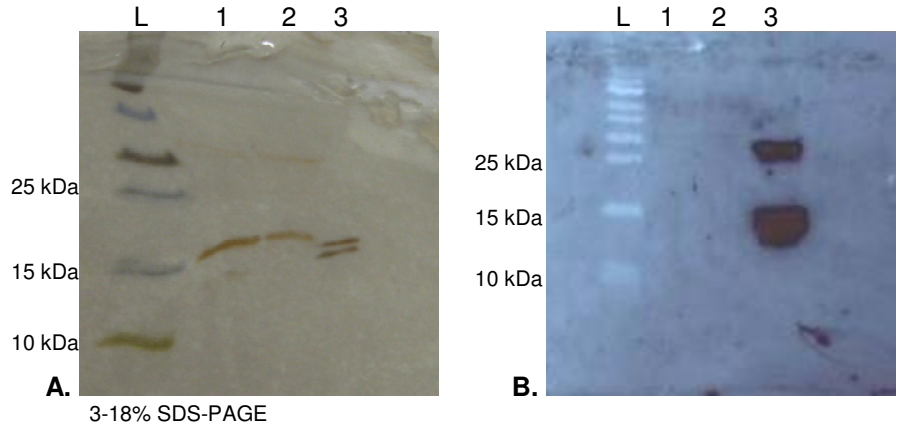


Figure 4.4. His-tagged EsxH was not detected in *M. smegmatis* WCLs. Although SDS-PAGE and silver staining detected proteins of the appropriate size for EsxH in the WCLs of both *M. smegmatis* WT and ESX-3_{ms} KO containing p19Kpro_EsxGH-His (A), this protein could not be verified by Western blotting with antibodies against EsxH (B) or the His-tag (not shown). L. PageRuler Prestained Protein Ladder (Fermentas), 1. *M.smegmatis* WT + EsxGH-His, 2. *M.smegmatis* ESX-3_{ms} KO + EsxGH-His, 3. TB10.4 positive control.

4.5. Discussion

This study aimed to determine whether the *M. smegmatis* ESX-3 secretion machinery is able to secrete *M. tuberculosis* EsxH protein, and thereby begin to establish whether the ESX-3 secretion system is functionally conserved between the two species. ESX-3 is essential in *M. tuberculosis* and cannot be deleted from the organism (Sassetti *et al.,* 2003), therefore *M. smegmatis* and the *M. smegmatis* ESX-3_{ms} KO are useful for the investigation of EsxH secretion by ESX-3. In addition the ease of manipulation of *M. smegmatis* and its rapid growth make it more convenient for analyses. By investigating the differential secretion of *M. tuberculosis* EsxH by *M. smegmatis* WT, ESX-3_{ms} KO and strains complemented with the *M. tuberculosis* ESX-3 it could be determined whether *M. smegmatis* is able to secrete *M. tuberculosis* EsxH, whether the ESX-3 secretion system is responsible for the secretion of EsxH in *M. smegmatis,* and whether introduction of the *M. tuberculosis* ESX-3 secretion system in *M. smegmatis* can complement an absence of EsxH secretion from this species.

EsxG and EsxH are coexpressed *in vivo* and interact to form a complex which is secreted (Okkels and Andersen, 2004). Therefore we coexpressed EsxG and EsxH in *M. smegmatis* to investigate the involvement of ESX-3 in EsxG-EsxH secretion by establishing the presence of EsxH (TB10.4) in the whole cell lysates and culture filtrates of *M. smegmatis* strains of various genotypes. However no EsxH could be detected in any of the samples, probably due to insufficient EsxH expression. Therefore we were unable to determine the ability of *M. smegmatis* ESX-3_{ms} to secrete *M. tuberculosis* EsxH. Further investigations were not possible due to time constraints, however expression of EsxG and EsxH from an inducible vector may enable more effective detection EsxH and further analysis of its secretion. This will be investigated in further studies.

Understanding the level of conservation of function of the ESX-3 between mycobacterial species, and its role in EsxH secretion may assist in establishing the *M. smegmatis* ESX-3_{ms} KO as a useful model for the investigation of the functions of ESX-3, and to establish its essential role in *M. tuberculosis.* As ESX-3 is essential for *M. tuberculosis* survival and EsxH, which is highly expressed under low iron and zinc conditions (Rodriguez *et al.,* 2002; Maciag *et al.,* 2007), is a potent antigen (Skjot *et al*., 2000), this secretion system is an important topic for tuberculosis research and a greater understanding of which may lead to important advances in tuberculosis prevention and treatment.

CHAPTER 5

Conclusion and Future directions

Conclusion

The *M. tuberculosis* ESAT-6 gene cluster region 3 encodes the machinery of a dedicated secretion system, ESX-3. This secretion system is essential for the survival of *M. tuberculosis*, therefore understanding its structure, function and substrates may lead to the identification of novel drug targets for the treatment of tuberculosis.

In this study we have implemented a novel technique, Mycobacterial – Protein Fragment Complementation (Singh *et al.,* 2006), to identify novel protein-protein interactions between the components of the ESX-3 secretion machinery. We identified several putative interactions and proposed a model of the mechanism of the ESX-3 secretion machinery based on these interactions, the putative functions of the proteins and the current knowledge of the ESX secretion systems. The Esx, PE and PPE proteins encoded in the ESAT-6 gene clusters are predicted to be secreted substrates of the ESX secretion machineries. We have shown that the EsxG-EsxH and PE5-PPE4 complexes have almost identical interactomes and are therefore probably secreted via the same secretion mechanism. In addition, the Esx proteins duplicated from those in the ESAT-6 gene cluster region 3 interact with EsxG and EsxH, and themselves, and therefore probably form additional substrates of the ESX-3 secretion machinery.

We have also created a knock-out of the ESX-3_{ms} secretion system in *M. smegmatis*, which will enable further investigation into the functions, substrates and mechanisms of this secretion system. As ESX-3 is essential in *M. tuberculosis*, the ESX-3_{ms} knock-out of *M. smegmatis* is a valuable tool for the investigation of this secretion system.

We attempted to investigate whether *M. smegmatis* is able to secrete *M. tuberculosis* EsxH and to assess the role of ESX-3_{ms} in the secretion thereof, by comparing EsxH secretion by *M. smegmatis* WT and the *M. smegmatis* ESX-3_{ms} KO. Unfortunately we were unable to detect any EsxH in these samples, probably due to inefficient expression of the proteins from the specific mycobacterial expression vectors that were used in this study. Understanding the conservation of function and mechanism of secretion of ESX-3 between *M. tuberculosis* and *M. smegmatis* would assist in validating the extrapolation of any findings regarding ESX-3, from studies in *M. smegmatis* to the *M. tuberculosis* ESX-3 secretion system and also confirm the role of ESX-3 in EsxH secretion.

Expression of the ESX-3 secretion machinery is upregulated under low iron and low zinc conditions (Rodriguez *et al.,* 2002; Maciag *et al.,* 2007). Mycobacteria employ two classes of siderophores for iron uptake; namely exochelins and mycobactins. *M. smegmatis* and other saprophytic mycobacteria produce both exochelins and mycobactins, however *M. tuberculosis* and other pathogenic bacteria produce only siderophores of the mycobactin class. A recent study by Siegrist *et al.* (2009) showed that ESX-3 is required for mycobactin utilisation in mycobacteria, allowing the uptake of iron from these scavenger molecules. This explains the disparity in the essentiality of ESX-3 in *M. smegmatis* and *M. tuberculosis*; while *M. smegmatis* can harvest iron from exochelins in the absence of ESX-3, *M. tuberculosis* is unable to acquire the iron necessary for its survival. A conditional knock-out of ESX-3 in *M. tuberculosis* is able to survive in media with very high concentrations of zinc and iron, or with culture filtrate from wild-type *M. smegmatis* (Serafini *et al.,* 2009). This suggests that ESX-3 is involved in the secretion of factors which allow for the uptake of mycobactin-scavenged iron into the mycobacterial cell. During infection iron is in demand from both the host and pathogen. The host sequesters iron to limit the amount available to the pathogen. Therefore the iron available to the infecting mycobacteria is severely limited. ESX-3 expression is upregulated under iron-poor conditions, leading to the increased secretion of factors involved in iron acquisition from mycobactin. EsxH, secreted by ESX-3, is a potent immunogen, consistent with the theory that secretion for ESX-3 is upregulated under iron-poor conditions during infection. Indeed, Siegrist *et al.* 2009 showed that mycobacteria lacking ESX-3 are severely impaired during macrophage infection.

The ESX-3 secretion system appears to play an important role in the survival of pathogenic mycobacteria, especially under iron-limiting conditions, as experienced in the host. In addition, the secretion system is required for infection. Therefore understanding this secretion system, its functions and substrates is essential to fully understand mycobacterial pathogenicity. We have identified interactions between components and substrates of this secretion system. These will assist in understanding the structure and function of this secretion system. In addition the *M. smegmatis* ESX-3_{ms} knock-out will enable further studies to identify substrates and the specific functioning of the secretion system. This study has laid groundwork for future investigations into ESX-3, a better understanding of which may enable advances in the treatment and prevention of tuberculosis and hopefully assist in the eradication of the disease.

Future directions

The protein-protein interactions between the components of the ESX-3 secretion machinery, identified in this study, need to be confirmed and validated using other techniques, possibly including *in vivo* pull down assays and Blue Native PAGE. As part of an international collaboration, additional studies are underway to identify protein-protein interactions from within the other 4 ESX secretion systems. Comparisons between the interactomes of the 5 ESX secretion systems may highlight similarities and differences between them. Interactions between components of different ESX secretion systems may also be investigated and enhance our understanding of the degree, if any, of complementation between them.

Future studies aim to further understand these mycobacterial secretion systems, including

- To investigate the evolutionary history of the type VII ESX secretion systems (T7SSs) by determining the presence and studying the function of these systems in different mycobacterial and other actinobacterial species.

- To determine the regulation of the ESX secretion systems.

- To investigate the structure and secretion mechanism of the ESX secretion systems.

- To identify and investigate novel substrates of the ESX secretion systems and to study their functions in the human host infection process.

- To determine the role of the ESX systems in divalent-cation homeostasis and PE-PPE interactions in the mycobacteria.

It is envisaged that these studies will allow us to better understand the functions of the five immunologically important ESX secretion systems of *M. tuberculosis*, thereby providing targets and biomarkers for future diagnostic, drug and vaccine design and facilitate in the eradication of tuberculosis disease.

ADDENDUM A

Materials and Methods

1. Bacterial strains

E. coli JM109 (laboratory strain) and K12 ER2925 (NEB) and *M. smegmatis* mc^2155 (Snapper *et al.*, 1990) are used as described in each study (Table A1).

2. Media and culture conditions

Culture media are described in Table A2. Antibiotics and supplements are described in Table A3.

E. coli is cultured in liquid Luria-Bertani broth (LB) with shaking, and on LB agar plates, overnight at 37 ℃. Solid and liquid media are supplemented with antibiotics; ampicillin (50 ug/ml), kanamycin (50 ug/ml) and hygromycin (100 ug/ml); and solid media with 5-bromo-4-chloro-3-indolyl-β-galactoside (X-gal) and isopropyl-β-D-thiogalactopyranoside (IPTG); as appropriate.

M. smegmatis is routinely grown in Middlebrook 7H9 Broth with shaking, and on BBL™ Seven H11 Agar Base plates, at 37 ℃ for 2-3 days unless otherwise described. For knock-out construction *M. smegmatis* is grown in LB with 0.1% Tween-80 and on LB agar, and for secretion studies *M. smegmatis* is grown in Kirchener's broth. Kanamycin (25 ug/ml), hygromycin (50 ug/ml) and trimethoprim (15-50 ug/ml), sucrose (5%) and X-gal are added as appropriate.

Bacterial strains used in this study are stored in 40% glycerol at -80 ℃.

3. Ziehl-Neelsen Staining

M. smegmatis cultures are screened for contamination using Ziehl-Neelsen (ZN) staining according to standard protocol as outlined by Kent and Kubica (1985). Briefly, aliquots of *M. smegmatis* cultures are heat-fixed to microscope slides. Slides are flooded with ZN Carbol Fuschin, heated intermittently with a flame until steaming, and allowed to stand for 5 minutes. Slides are rinsed with water, decolourised with 5% acid-alcohol solution for 2 minutes and rinsed with water. Slides are then counterstained with Methylene Blue for 1-2 minutes, rinsed with water and allowed to air dry. Slides are read using a light microscope under the 100X (oil immersion) lens. Uncontaminated *M. smegmatis* cultures will contain only acid fast bacilli (pink rods).

Table A1. Bacterial Strains

Strain	Genotype	Features	Reference/Source
E. coli JM109	*recA1, endA1, gyrA96, thi, hsdR17, supE44, relA1, Δ(lac-proAB)/F' [traD36, proAB+, lacIq, lacZΔM15*	Chloramphenicol resistance Methylation: dam+, dcm+ Blue-white selection +	Laboratory strain
E. coli K12 ER2925	*ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 R(zgb210::Tn10)TetS endA1 rpsL136 dam13::Tn9 xylA-5 mtl-1 thi-1 mcrB1 hsdR2*	Tetracycline resistance Methylation: dam- , dcm- Blue-white selection -	New England Biolabs
M. smegmatis mc²155	*ept-1*	Efficient plasmid transformation mutant of *M. smegmatis* mc²6 Mycobacterial model organism	Snapper *et al.* 1990

Table A2. Culture media

Medium	Composition	Source
Luria-Bertani broth (LB)	1% Tryptone, 0.5% Sodium chloride, 0.5% Yeast extract	
Luria-Bertani agar (LB agar)	1% Tryptone, 0.5% Sodium chloride, 0.5% Yeast extract, 1.2% bacteriological agar	
SOB	2% Tryptone, 0.5% Sodium chloride, 0.5% Yeast extract, 2.5 mM Potassium chloride,	
SOC	SOC supplemented with 20 mM glucose and 10 mM Magnesium chloride after autoclaving	
Middlebrook 7H9 Broth (7H9)	0.47% Middlebrook 7H9 Broth powder, supplemented with 0.5% glycerol, 0.5% glucose, 0.2% Tween-80 after autoclaving	Becton Dickinson (BD)
BBL™ Seven H11 Agar Base (7H11)	1.9% BBL™ Seven H11 Agar Base powder, supplemented with 0.5% glycerol, 0.5% glucose, 0.2% Tween-80 after autoclaving	Becton Dickinson (BD)
Kirchener's Broth	0.5% asparagine, 0.3% DiSodium hydrogen phosphate, 0.4% Potassium dihydrogen phosphate, 0.25% Trisodium citrate, 0.107% Magnesium sulphate, 2% glycerol	

All media was prepared in deionised water and autoclaved at 121 ℃ for 20 minutes. Concentrations of all dry components are given as % m/v, and liquids as % v/v.

Table A3. Antibiotics and supplements

	Stock concentration	Solvent	Sterilisation	Storage	Supplier	Working concentration	
						E. coli	*M. smegmatis*
Ampicillin (amp)	50 mg/ml	Deionised water	filtered	-20 ℃	Roche	50 ug/ml	n/a
Chloramphenicol (cam)	34 mg/ml	Ethanol		4 ℃	Roche	25 ug/ml	n/a
Hygromycin* (hyg)	50 mg/ml	Phosphate buffered saline (PBS)	filtered	4 ℃	Invitrogen	100 ug/ml	50 ug/ml
Kanamycin* (kan)	50 mg/ml	0.9% NaCl solution	filtered	4 ℃	Sigma	50 ug/ml	25 ug/ml
Tetracycline (tet)	5 mg/ml	Ethanol		-20 ℃	Sigma	50 ug/ml	n/a
Trimethoprim (trim)	50 mg/ml	DiMethyl Sulfoxide (DMSO)	filtered	use fresh	Sigma	n/a	15-50 ug/ml
5-bromo-4-chloro-3-indolyl-β-galactoside (X-gal)	20 mg/ml	N,N'-dimethyl formamide		-20 ℃, in foil	Roche	100 ul spread on 90 mm plate	
Isopropyl-β-D-thiogalactopyranoside (IPTG)	0.1M	Deionised water	filtered	-20 ℃	Biosolve	100 ul spread on 90 mm plate	
Potassium Chloride (KCl)	1M	Deionised water	autoclaved, 121℃, 20min	RT	Merck		
Magnesium Chloride* ($MgCl_2$)	1M	Deionised water	filtered	RT	Sigma		
Glucose	1M	Deionised water	filtered	-20 ℃	Merck		
Glucose	50% v/v	Deionised water	filtered	4 ℃	Merck		0.5% v/v
Glycerol	50% v/v	Deionised water	filtered	4 ℃	Merck		0.5% v/v
Polyoxyethylene-sorbitan monooleate (Tween 80)	20% v/v	Deionised water	filtered	4 ℃	Sigma		0.2% v/v
Sucrose	50% v/v	Deionised water	filtered	4 ℃	Merck		5% v/v

*Supplements and antibiotics purchased in solution

4. DNA manipulation

4.1. DNA and protein sequence analyses

All *M. tuberculosis* DNA and protein sequence information is obtained from the *M. tuberculosis* H37Rv genome sequence database (Tuberculist, http://genolist.pasteur.fr/TubercuList/index.html). The *M. smegmatis* DNA sequence is obtained from the JVCI CMR *Mycobacterium smegmatis* MC2 Genome Page (http://cmr.jcvi.org/tigr-scripts/CMR/GenomePage.cgi?org=gms).

4.2. Primer design

Primers are designed using Primer Premier 5.0 (PREMIER Biosoft International) and obtained from Integrated DNA Technologies (IDT, through Whitehead Scientific, www.whitesci.co.za). Appropriate restriction endonuclease cutting sites are incorporated as necessary to facilitate in-frame cloning procedures. Primer sequences and details are given in each study.

4.3. PCR amplification

FastStart Taq DNA Polymerase (Roche) is used for PCR amplification using the GC-rich buffer, as described by the manufacturer. Annealing temperatures (Tm) appropriate for each primer set are given in each study and elongation times calculated at 1 minute per 1kb. Thirty-five cycles are used unless otherwise stated.

4.4. Restriction Digestion

Plasmid DNA is digested using restriction endonucleases as per manufacturers instructions, using the appropriate buffers. 2-5 ug of plasmid DNA is digested using 10U of enzyme in a total volume of 10-20 ul for 2-5 hours at 37 ℃. All restriction endonucleases are obtained from New England Biolabs (NEB), Promega or Fermentas.

4.5. Agarose gel electrophoresis

PCR products and restriction digests are separated by electrophoresis at 80V on 1.0 to 1.5% agarose gels in TAE buffer (40 mM Tris, 20 mM Acetic acid, 1 mM Ethylenediamine tetraacetic acid (EDTA)) with

ethidium bromide and visualized using UV light. The GeneRuler™ 100bp DNA Ladder Plus, ready-to-use (Fermentas) is used to determine band size. When necessary the resultant bands are excised from the gel and purified.

4.6. DNA clean-up

The Wizard® SV Gel and PCR Clean-up System (Promega) is used to purify DNA embedded in agarose gel slices and directly from a PCR or restriction endonuclease reaction.

4.7. DNA quantification

DNA is quantified using the NanoDrop® ND1000 Spectrophotometer.

4.8. DNA sequencing

Automated sequencing is done by the Central Analytical Facility (CAF) at Stellenbosch University, using the ABI 3130XL Genetic Analyzer (Applied Biosystems).

4.9. Cloning vectors

All inserts are cloned into the commercial T-vector, pGem-T-easy (Promega), and subcloned into the relevant expression and suicide vectors, as described in each study.

4.10. Dephosphorylation

Linearized vectors are dephosphorylated using Shrimp Alkaline Phosphatase (SAP, Roche) to prevent self-ligation of the digested plasmids. One unit of SAP is added to digested vector DNA with the appropriate buffers, according to the manufacturer's instructions. The reaction is incubated for 10 minutes at 37 °C, followed by deactivation for 15 minutes at 65 °C.

4.11. Ligations

PCR and restriction digest products are ligated into appropriate vectors using T4 DNA Ligase (Promega), according to manufacturer's instructions. Inserts are ligated into 100 ng of vector at an insert:vector molar

ratio of 3:1 using 3u of T4 DNA Ligase, and incubated at 4 ℃ overnight. Ligation reactions are deactivated at 70 ℃ for 10 minutes prior to electrotransformation.

4.12. Transformations

All transformations of *E. coli* and *M. smegmatis* are by electroporation of electrocompetent cells, as described below, using the Gene Pulser™ (Bio-Rad) in 0.2 cm Gene Pulser® Cuvettes (Bio-Rad).

4.13. Colony PCR

All transformants are confirmed by colony PCR using 1 ul culture in a 12.5 ul PCR reaction as described under PCR amplification.

4.14. Miniprep plasmid isolation

Plasmid DNA is isolated from *E. coli* using the Wizard® *Plus* SV Minipreps DNA Purification System (Promega), according to manufacturers instructions, and eluted in 50 ul nuclease-free water (Promega).

5. Cloning

Inserts are amplified by PCR using specific PCR primers. The PCR products are separated by electrophoresis, purified and ligated into pGem-T-Easy according to manufacturer's instructions. The ligation mixture is transformed into competent *E. coli* and selected for on appropriate LB agar. Colonies are picked and inoculated into LB media. Colony PCRs are done to confirm the insert and the construct was isolated. The resultant constructs are sequenced using Sp6 and T7 primers to confirm the sequence of the inserts.

The constructs are digested using the primer-specific restriction endonucleases and the inserts separated by electrophoresis, cut from the gel and purified. The purified inserts are ligated into the pre-digested expression or suicide vectors, as appropriate, and transformed into *E. coli.* Transformed colonies are picked, colony PCR done and the constructs isolated and sequenced to verify the sequence and frame of relevant inserts.

6. Preparation of electrocompetent cells

6.1. Electrocompetent *E. coli*

E. coli JM109 or K12 ER2529 cells are inoculated into 50 ml LB containing tetracycline (50 ug/ml) or chloramphenicol (25 ug/ml), respectively, and shaken overnight at 37ºC. Overnight cultures are inoculated into fresh antibiotic-containing medium, at a dilution of 1:100, and grown to OD_{600} = 0.7 at 37ºC, with shaking. Cultures are placed on ice and all further steps performed under ice-cold conditions. Cells are harvested by centrifugation at 5000 rpm for 10 minutes at 4ºC. The cell pellet is resuspended in an equal culture-volume of ice-cold 10% glycerol and centrifuged at 5000 rpm for 10 minutes at 4ºC. The wash step is repeated and the cells pooled in a 50 ml falcon tube and centrifuged at 4000 rpm for 10 minutes at 4ºC. The cells are resuspended in ice-cold 10% glycerol using a volume of 2 ml per liter of culture. Aliquots of 100 ul are prepared, frozen in liquid nitrogen and stored at -80ºC.

6.2. Electrocompetent *M. smegmatis*

M. smegmatis mc^2155 cells are inoculated into Middlebrook 7H9 medium and grown to OD_{600} = 0.5, at 37ºC with shaking. Cultures are incubated on ice for 1 hour and all further steps performed under ice-cold conditions. Cells are pelleted by centrifugation at 4000 rpm for 10 minutes at 4ºC, resuspended in ice-cold 10% glycerol equal to the culture volume and pelleted. The wash step is repeated and the cells resuspended in ice-cold 10% glycerol using a volume of 2 ml per 100 ml of culture. Electrocompetent *M. smegmatis* is freshly prepared before use.

7. Electroporation

7.1. *E. coli* electroporation

E. coli is transformed at 2.5 kV, 25 µF, 125µ Fd, 200 Ω using 2 ul of a 10 ul ligation reaction, or 1 ug of plasmid DNA, in 45 ul of electrocompetent cells. One milliliter of SOC is added to the transformed cells which are expressed for 30-60 minutes at 37 ºC before plating on appropriate solid media.

7.2. *M. smegmatis* electroporation

One microgram of purified plasmid DNA is transformed into 200 ul of electrocompetent *M. smegmatis* at 2.5 kV, 25 µF, 125 µFd, 1000 Ω followed by the addition of 1 ml 7H9 medium and expression for 3 hours at 37 ℃ before plating on appropriate solid medium.

8. Protein analyses

8.1. Culture fractionation

M. smegmatis cultures are separated into cell pellet and culture supernatant by centrifugation at 4000 rpm for 15 minutes at 4 ℃. All subsequent steps are performed on ice.

8.1.1. Culture filtrate protein precipitation

The culture supernatant is filtered consecutively through 0.45 um and 0.22 um syringe filters. Four volumes of ice cold acetone is added to the culture filtrate, and incubated at -20 ℃ overnight. The culture filtrate proteins are precipitated by centrifugation at 14 000 rpm for 15 minutes at 4 ℃. The supernatant is discarded and the protein pellet stored at -80 ℃. The pellet is resuspended in phosphate-buffered saline (PBS) prior to use.

8.1.2. Whole cell lysate

The cell pellet is resuspended in 1 ml ice cold PBS and transferred to a 2 ml ribolyser tube. The sample is centrifuged at 14 000 rpm at 4 ℃ for 10 minutes, and the supernatant discarded. A volume of 0.1 mm silica beads equal to the volume of the cell pellet, and 500ul ice cold PBS are added to the tube. The cells are mechanically lysed by ribolysing in a BIO 101/Savant FastPrep FP120, 6 times for 20 seconds at maximum speed, with 1 minute intervals on ice. The samples are incubated on ice for 5 minutes and centrifuged for 10 minutes at 14 000 rpm at 4 ℃. The supernatant, containing the whole cell lysate, is filtered through a 0.22 um syringe filter and stored at -80 ℃.

8.2. Purification of His-tagged proteins from Whole Cell Lysate

The MagneHis™ Protein Purification System (Promega) is used, essentially as described by the manufacturers, to purify the His-tagged proteins from *M. smegmatis* cells. Cultures are fractionated and the cell lysate prepared as described in 8.1 in MagneHis Binding/Wash Buffer. Two microliters of preprepared DNase I is added to each sample and incubated, with shaking, for 30 minutes. The sample is centrifuged at 4000 rpm for 10 minutes and the supernatant removed. NaCl is added to each sample at final concentration of 500 mM. Thirty microliters of MagneHis Ni-Particles are added to each 1 ml sample and mixed by inverting 10 times. The samples are incubated at room temperature for 5 minutes to allow protein binding to the Nickel beads. The samples are placed in a magnetic stand for 30 seconds to capture the beads, and the supernatant removed. The beads are washed three times with MagneHis Binding/Wash Buffer (with 500 mM NaCl): 150 ul of Wash buffer is added to the sample and mixed by pipetting, the sample is placed in the magnetic stand for 30 seconds and the supernatant removed. The His-tagged proteins are eluted by adding 100µl of MagneHis Elution Buffer to the beads, pipetting to mix, and incubating at room temperature for 2 minutes. The tubes are placed in the magnetic stand, to capture the beads, and the supernatant, containing His-tagged proteins removed and stored at -20 ℃ for future applications.

8.3. Protein concentration determination

The protein concentrations of samples are determined spectrophotometrically using the Bio-Rad Protein Assay with the 1 x Quick Start™ Bradford Dye Reagent (Bio-rad), according to standard microtiter plate procedure, as described by the manufacturer,. The Quick Start™ Bovine Serum Albumin Standard Set (BSA concentration range 0.125 mg/ml – 2 mg/ml) is used to generate a standard curve.

Protein samples are measured in undiluted, 1:10 and 1:100 dilutions, in duplicate. Standards are measured in triplicate. Ten microliters of each sample and standard are pipetted into separate microtiter wells. Two hundred microliters of Dye Reagent is added to each well, and mixed. The samples are incubated at room temperature for 5 minutes. Absorbance is measured at 595 nm (read within 1 hour).

A standard curve is plotted from the BSA standard absorbance readings, and sample concentrations determined from the standard curve.

8.4. SDS-PAGE

All reagents and buffers used for SDS-PAGE are described in Table A4 and A5.

8.4.1. Sample preparation

An equal volume of 2 x SDS Sample Buffer was added to each protein sample (maximum total volume per well is 20 ul). The sample is mixed and incubated at 100 ℃ for 5 minutes.

8.4.2. Gel preparation

The gel apparatus is prepared according to manufacturer's instructions. The apparatus is filled with water to check for leakage.

The separating gel is prepared according to Table A6 and poured into the gel apparatus. Isopropanol is added to the surface of the gel to remove bubbles and level the surface of the gel. Allow 20 minutes for the gel to set, pour off the isopropanol and rinse the gel surface with Stacking buffer.

Prepare stacking gel according to Table A6 and pour over separating gel until gel is overfull. Insert comb. Allow gel to set for 20 minutes.

8.4.3. Electrophoresis

Remove gel from setting apparatus and place in running chamber. Pour running buffer into upper and lower chambers to cover top and bottom of gel, and both electrodes. Remove comb. Load 15-20 ul of samples into wells, including molecular weight marker.

Electrophorese at 100 V until samples enter separating gel, increase voltage to 150 V and run until the dye front reaches the end of the gel.

Table A4. SDS-PAGE reagents

Reagent		concentration	Diluent	Storage	Source
20% SDS	Sodium dodecyl sulphate	20% m/v	Water	RT	Sigma
10% SDS	Sodium dodecyl sulphate	10% m/v	Water	RT	Sigma
EDTA	Ethylenediaminetetraacetic acid	0.5 M	Water	RT	Fluka
Bromophenol blue		0.75% m/v	Water	RT	Fluka
APS	Ammonium persulphate	10% m/v	Water	4 ℃	Sigma
Acrylamide/bisacrylamide		30:0.8% m/v	Water	4 ℃, in dark container	Bio-rad
TEMED	N,N,N',N'-Tetramethylethylenediamine			RT, in dark container	Sigma
β-Mercaptoethanol				RT, in dark container	Merck
Glycine				RT	Merck
Tris	tris(hydroxymethyl)aminomethane			RT	Merck
PageRuler™ Prestained Protein Ladder				-20 ℃	Fermentas

Table A5. SDS-PAGE buffers

Buffer	Component	Volume/concentration	Diluent	Storage	pH
Separating buffer	Tris	1.5 M	Water	RT	pH 8.8 (adjust with HCl)
Stacking buffer	Tris	1.0 M	Water	RT	pH 6.8 (adjust with HCl)
2x SDS Sample buffer	Stacking buffer	3.4 ml		-20 ℃ or 4 ℃	
	Glycerol	2 ml			
	20% SDS	3 ml			
	Bromophenol blue	500 ul			
	EDTA	200 ul			
	β-Mercaptoethanol	1 ml			
Running buffer	Tris	25 mM	Water	4 ℃, or RT	
(do not adjust pH)	Glycine	192 mM			
	SDS	0.1% m/v			

Table A6. SDS-PAGE gel composition

Component	18% Separating gel (2 x 0.75 mm mini-slab gels)	Stacking gel (4 x 0.75 mm mini-slab gels)
Acrylamide/bisacrylamide	6.0 ml	1.3 ml
Separating buffer	2.5 ml	
Stacking buffer		1.25 ml
10% SDS	0.1 ml	0.1 ml
TEMED	10 ul	20 ul
APS	50 ul	50 ul
Water	fill to 10 ml	fill to 10 ml

8.5. Silver staining

Silver staining was done as described by Gromova *et al.* (2006). Buffers used for silver staining are described in Table A7.

Soak SDS-PAGE gel in fixation solution for 2 hours, or overnight. Discard fixation solution and wash the gel in washing solution for 20 minutes, replacing the washing solution 3 times. Discard washing solution and shake gently in sensitizing solution for 2 minutes. Discard sensitizing solution and wash gel twice with pre-chilled deionised water for one minute. Discard water and add cold silver nitrate staining solution (add the solution to the corner of the tray and not directly onto the gel). Shake for 1 hour at 4 °C. Pour off the staining solution and rinse the gel twice with deionised water for 1 minute. Rinse gel in developing solution, discard and develop protein image in fresh developing solution for 2 - 5 minutes. Stop the reaction by adding terminating solution directly to the gel in the developing solution. Gently shake the gel in solution for about 10 minutes until the solution stops bubbling.

8.6. Western Blotting

Western Blotting is done as described by Gey van Pittius (2005). Buffers used in Western Blotting are described in Table A8.

8.6.1. Protein transfer

Presoak 2 western blot pads, 2 Whatmann 3M sheets and 1 HybondC-extra nitrocellulose or Hybond-P PVDF membrane in transfer buffer. When electrophoresis is completed, stack the following items onto the black side of the blotting apparatus sandwich, in transfer buffer: pad, Whatmann sheet, gel, membrane, Whatmann sheet, pad. Close sandwich and insert into Western blot apparatus with an ice pack and fill with cold transfer buffer. Add magnetic stirrer bar to chamber and place on magnetic stirrer in coldroom. Apply 100 V for 1 hour. Buffer becomes milky with air bubbles when current is flowing.

Dismantle apparatus and check transfer by staining with Ponceau Red dye. The membrane is rinsed twice in distilled water, Ponceau Red dye added, and gently rotated for 5 minutes. The dye is retained for future

use, the membrane rinsed in distilled water and the protein bands observed. The dye is rinsed from the membrane with distilled water.

The membrane is blocked overnight in 10% Blocking buffer.

8.6.2. Western detection

The blocked membrane is rinsed twice in TBS-T, and washed 3 times for 10 minutes in TBS-T buffer with shaking. Five milliliters of primary antibody, at the appropriate dilution, is added to the membrane and rotated at room temperature for 3 hours (Primary antibody may be retained and reused several times). The membrane is rinsed twice in TBS-T, and washed 3 times for 10 minutes in TBS-T buffer with shaking. The membrane is then probed with 5 ml secondary antibody for 1 hour at room temperature, washed as before and placed in a sealable plastic bag.

The membrane is incubated in ECL detection fluid (ECL Plus Western Blotting Detection System, Amersham™) according to manufacturers instructions. The detection fluid is removed, the membrane attached to a piece of Whatmann paper and placed in a cassette. Signal is detected for 1-10 minutes on with autoradiographic film.

Membranes can be stripped once in preheated stripping buffer in a 50 ℃ shaking water bath for 30 minutes, washed twice in TBS-T at room temperature and reblocked in blocking buffer, followed by Western detection with an additional probe.

Table A7. Silver staining solutions

Solution	Component	Source	Diluent	storage
Fixation solution	50% ethanol	Merck	water	RT
	12% acetic acid	Merck		
	0.05% formalin	Merck		
Washing solution	20% ethanol	Merck	water	RT
Sensitizing solution	0.02% m/v sodium thiosulfate	Sigma	water	RT
Staining solution	0.2% m/v silver nitrate	Sigma	water	prepare fresh
	0.076%formalin	Merck		precool to 4 ℃ before use
Developing solution	6% m/v sodium carbonate	Merck	water	RT
	0.0004% sodium thiosulphate	Sigma		
	0.05% formalin	Merck		
Terminating solution	12% acetic acid	Merck	water	RT
Drying solution	20% ethanol	Merck	water	RT

Table A8. Western blotting buffers

Buffer	Composition	Diluent	pH	Storage
Transfer buffer	25 mM Tris	Water	pH should be 8.3; do not adjust	4 ℃
	192 mM Glycine			
	20% Methanol			
TBS-T	20 mM Tris (pH 7.6)	Water		RT
	137 mM NaCl			
	0.1% Tween-20			
Blocking buffer	10% m/v fat free milk powder	TBS-T		RT
	1% m/v Bovine serum albumin			
	10 mM Azide			
Stripping buffer	100 mM β-Mercaptoethanol	Water		RT
	2% m/v SDS			
	62.5 mM Tris (pH6.7)			

REFERENCE LIST

Abdallah,A.M., Verboom,T., Hannes,F., Safi,M., Strong,M., Eisenberg,D. *et al.* (2006) A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. *Mol Microbiol* **62**: 667-679.

Abdallah,A.M., Gey van Pittius,N.C., Champion,P.A., Cox,J., Luirink,J., Vandenbroucke-Grauls,C.M. *et al.* (2007) Type VII secretion--mycobacteria show the way. *Nat Rev Microbiol* **5**: 883-891.

Abdallah,A.M., Savage,N.D., van Zon,M., Wilson,L., Vandenbroucke-Grauls,C.M., van der Wel,N.N. *et al.* (2008) The ESX-5 secretion system of *Mycobacterium marinum* modulates the macrophage response. *J Immunol* **181**: 7166-7175.

Abdallah,A.M., Verboom,T., Weerdenburg,E.M., Gey van Pittius,N.C., Mahasha,P.W., Jimenez,C. *et al.* (2009) PPE and PE_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. *Mol Microbiol*. **73**: 329-340.

Abou-Zeid,C., Garbe,T., Lathigra,R., Wiker,H.G., Harboe,M., Rook,G.A., and Young,D.B. (1991) Genetic and immunological analysis of *Mycobacterium tuberculosis* fibronectin-binding proteins. *Infect Immun* **59**: 2712-2718.

Adindla,S., and Guruprasad,L. (2003) Sequence analysis corresponding to the PPE and PE proteins in *Mycobacterium tuberculosis* and other genomes. *J Biosci* **28**: 169-179.

Andersen,P., Andersen,A.B., Sorensen,A.L., and Nagai,S. (1995) Recall of long-lived immunity to *Mycobacterium tuberculosis* infection in mice. *J Immunol* **154**: 3359-3372.

Banu,S., Honore,N., Saint-Joanis,B., Philpott,D., Prevost,M.C., and Cole,S.T. (2002) Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol Microbiol* **44**: 9-19.

Bayan,N., Houssin,C., Chami,M., and Leblon,G. (2003) Mycomembrane and S-layer: two important structures of *Corynebacterium glutamicum* cell envelope with promising biotechnology applications. *J Biotechnol* **104**: 55-67.

Begg,K.J., Dewar,S.J. and Donachie,W.D. (1995) A new *Escherichia coli* cell division gene, *ftsK*. *J Bacteriol* **177**: 6211-6222.

Behr,M.A., Wilson,M.A., Gill,W.P., Salamon,H., Schoolnik,G.K., Rane,S., and Small,P.M. (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**: 1520-1523.

Berthet,F.X., Rasmussen,P.B., Rosenkrands,I., Andersen,P., and Gicquel,B. (1998) A *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10). *Microbiology* **144**: 3195-3203.

Bitter,W., Houben,E.N.G., Bottai,D., Brodin,P., Brown,E.J., Cox,J.S. *et al.* (2009) Systematic Genetic Nomenclature for Type VII Secretion Systems. *Plos Pathog* **5**: e1000507.doi:10.1371/journal.ppat.1000507.

Bönemann,G., Pietrosiuk,A., Diemand,A., Zentgraf,H. and Mogk,A. (2009) Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. *EMBO J* **28**: 315-325.

Brennan,M.J., Delogu,G., Chen,Y., Bardarov,S., Kriakov,J., Alavi,M., and Jacobs,W.R., Jr. (2001) Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. *Infect Immun* **69**: 7326-7333.

Brodin,P., Eiglmeier,K., Marmiesse,M., Billault,A., Garnier,T., Niemann,S. *et al.* (2002) Bacterial artificial chromosome-based comparative genomic analysis identifies *Mycobacterium microti* as a natural ESAT-6 deletion mutant. *Infect Immun* **70**: 5568-5578.

Brodin,P., Majlessi,L., Marsollier,L., de Jonge,M.I., Bottai,D., Demangel,C. *et al.* (2006) Dissection of ESAT-6 system 1 of *Mycobacterium tuberculosis* and impact on immunogenicity and virulence. *Infect Immun* **74**: 88-98.

Brosch,R., Gordon,S.V., Billault,A., Garnier,T., Eiglmeier,K., Soravito,C., Barrell,B.G. and Cole,S.T. (1998) Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. *Infect Immun* **66**: 2221-2229.

Brosch,R., Philipp,W.J., Stavropoulos,E., Colston,M.J., Cole,S.T. and Gordon,S.V. (1999) Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra strain. *Infect Immun* **67**: 5768-5774.

Brosch,R., Gordon,S.V., Buchrieser,C., Pym,A.S., Garnier,T., and Cole,S.T. (2000a) Comparative genomics uncovers large tandem chromosomal duplications in *Mycobacterium bovis* BCG Pasteur. *Yeast* **17**: 111-123.

Brosch,R., Gordon,S.V., Pym,A., Eiglmeier,K., Garnier,T., and Cole,S.T. (2000b) Comparative genomics of the mycobacteria. *Int J Med Microbiol* **290**: 143-152.

Brown,G.D., Dave,J.A., Gey van Pittius,N.C., Stevens,L., Ehlers,M.R., and Beyers,A.D. (2000) The mycosins of *Mycobacterium tuberculosis* H37Rv: a family of subtilisin-like serine proteases. *Gene* **254**: 147-155.

Burts,M.L., Williams,W.A., DeBord,K. and Missiakas,D.M. (2008) EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections. *PNAS* **102:** 1169-1174.

Callahan,B., Nguyen,K., Collins,A., Valdes,K., Caplow,M., Crossman,D.K., Steyn,A.J., Eisele,L., Derbyshire,K.M. (2009) Conservation of structure and protein-protein interactions mediated by the secreted mycobacterial proteins EsxA, EsxB, and EspA. *J Bacteriol* [Epub ahead of print]

Camus,J.C., Pryor,M.J., Medigue,C., and Cole,S.T. (2002) Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* **148**: 2967-2973.

Champion,P.A., Stanley,S.A., Champion,M.M., Brown,E.J., and Cox,J.S. (2006) C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*. *Science* **313**: 1632-1636.

Christie,P.J., Atmakuri,K., Krishnamoorthy,V., Jakubowski,S. and Cascales,E. (2005) Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu Rev Microbiol* **59**: 451-485.

Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C., Harris,D. *et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537-544.

Cole,S.T. (1999) Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. *FEBS Lett* **452**: 7-10.

Converse,S.E., and Cox,J.S. (2005) A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in *Mycobacterium smegmatis*. *J Bacteriol* **187**: 1238-1245.

Coros,A., Callahan,B., Battaglioli,E., and Derbyshire,K.M. (2008) The specialized secretory apparatus ESX-1 is essential for DNA transfer in *Mycobacterium smegmatis*. *Mol Microbiol* **69**: 794-808.

Dave,J.A., Gey van Pittius,N.C., Beyers,A.D., Ehlers,M.R., and Brown,G.D. (2002) Mycosin-1, a subtilisin-like serine protease of *Mycobacterium tuberculosis*, is cell wall-associated and expressed during infection of macrophages. *BMC Microbiol* **2**: 30.

Delogu,G., Pusceddu,C., Bua,A., Fadda,G., Brennan,M.J., and Zanetti,S. (2004) Rv1818c-encoded PE_PGRS protein of *Mycobacterium tuberculosis* is surface exposed and influences bacterial cell structure. *Mol Microbiol* **52**: 725-733.

Doran,T.J., Hodgson,A.L., Davies,J.K., and Radford,A.J. (1992) Characterisation of a novel repetitive DNA sequence from *Mycobacterium bovis*. *FEMS Microbiol Lett* **75**: 179-185.

Espitia,C., Laclette,J.P., Mondragon-Palomino,M., Amador,A., Campuzano,J., Martens,A. *et al.* (1999) The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins? *Microbiology* **145**: 3487-3495.

Euzeby,J.P (2009) List of Prokaryotic names with Standing in Nomenclature. http://www.bacterio.cict.fr/.

Flint,J.L., Kowalski,J.C., Karnati,P.K., and Derbyshire,K.M. (2004) The RD1 virulence locus of *Mycobacterium tuberculosis* regulates DNA transfer in *Mycobacterium smegmatis*. *Proc Natl Acad Sci U S A* **101**: 12598-12603.

Fortune,S.M., Jaeger,A., Sarracino,D.A., Chase,M.R., Sassetti,C.M., Sherman,D.R. *et al.* (2005) Mutually dependent secretion of proteins required for mycobacterial virulence. *Proc Natl Acad Sci U S A* **102**: 10676-10681.

Gao,L.Y., Guo,S., McLaughlin,B., Morisaki,H., Engel,J.N., and Brown,E.J. (2004) A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. *Mol Microbiol* **53**: 1677-1693.

Garufi,G., Butler,E. and Missiakas,D. (2008) ESAT-6-Like Protein Secretion in *Bacillus anthracis. J Bacteriol* **190**: 7004-7011.

Gelber,R.H. (1994) Chemotherapy of lepromatous leprosy: recent developments and prospects for the future. *Eur J Clin Microbiol Infect Dis* **13**: 942-952.

Gey van Pittius,N.C., Gamieldien,J., Hide,W., Brown,G.D., Siezen,R.J., and Beyers,A.D. (2001) The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. *Genome Biol* **2**: RESEARCH0044.

Gey van Pittius,N.C. (2005) Protein separation and detection – SDS PAGE and Western Blotting. HonsBScMedSci Practical 2005.

Gey van Pittius,N.C., Sampson,S.L., Lee,H., Kim,Y., van Helden,P.D., and Warren,R.M. (2006) Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. *BMC Evol Biol* **6**: 95.

Giuseppe Champion,P.A., Champion,M.M., Manzanillo,P., and Cox,J.S. (2009) ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria. *Mol Microbiol* **73**: 950-962.

Gordhan,B.G., and Parish,T. (2001) Gene Replacement using Pretreated DNA. In *Mycobacterium tuberculosis* Protocols. Eds. Parish,T. and Stoker N.G. (Humana Press, New Jersey) pp. 77-92.

Gordon,S.V., Brosch,R., Billault,A., Garnier,T., Eiglmeier,K., and Cole,S.T. (1999a) Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. *Mol Microbiol* **32**: 643-655.

Gordon,S.V., Eiglmeier,K., Brosch,R., Garnier,T., Honore,N., Barrell,B.G., and Cole,S.T. (1999b) Genomics of *Mycobacterium tuberculosis* and *Mycobacterium leprae*. In Mycobacteria:

molecular biology and virulence. Eds. Ratledge,C., and Dale,J. (Oxford: Blackwell Science Ltd) pp. 93-109.

Gromova,I. and Celis,J.E. (2006) Protein Detection in Gels by Silver Staining: A Procedure Compatible with Mass-Spectrometry. In Cell Biology: A Laboratory Handbook. Eds. Celis,J.E., Carter,N., Hunter,T., Simons,K., Small,J.V. and Shotton,D. (Elsevier. Academic Press).

Guinn,K.M., Hickey,M.J., Mathur,S.K., Zakel,K.L., Grotzke,J.E., Lewinsohn,D.M. *et al.* (2004) Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* **51**: 359-370.

Hermans,P.W., van,S.D., and van Embden,J.D. (1992) Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium gordonae*. *J Bacteriol* **174**: 4157-4165.

Hsu,T., Hingley-Wilson,S.M., Chen,B., Chen,M., Dai,A.Z., Morin,P.M. *et al.* (2003) The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc Natl Acad Sci U S A* **100**: 12420-12425.

Ize,B., and Palmer,T. (2006) Microbiology. Mycobacteria's export strategy. *Science* **313**: 1583-1584.

Kent,B. and Kubica,G. (1985). Public Health Mycobacteriology*:* a Guide for the Level II Laboratory. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control.

Koch,R. (1882) Die Aetiologie der Tuberkulose. *Berl Klin Wochensch* **19**: 221-230.

Le,M.V., Robreau,G., Borot,C., Guesdon,J.L., and Mahana,W. (2005) Expression, immunochemical characterization and localization of the *Mycobacterium tuberculosis* protein p27. *Tuberculosis (Edinb)* **85**: 213-219.

Lewis,K.N., Liao,R., Guinn,K.M., Hickey,M.J., Smith,S., Behr,M.A., and Sherman,D.R. (2003) Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guerin attenuation. *J Infect Dis* **187**: 117-123.

Li,Y., Miltner,E., Wu,M., Petrofsky,M., and Bermudez,L.E. (2005) A *Mycobacterium avium* PPE gene is associated with the ability of the bacterium to grow in macrophages and virulence in mice. *Cell Microbiol* **7**: 539-548.

Luthra,A., Mahmood,A., Arora,A., and Ramachandran,R. (2008) Characterization of Rv3868, an essential hypothetical protein of the ESX-1 secretion system in *Mycobacterium tuberculosis*. *J Biol Chem* **283**: 36532-36541.

MacCallum,P., Tolhurst,G., Buckle,G., and Sissons,H.A. (1948) A new mycobacterial infection in man. I. Clinical aspects. II. Experimental investigations in laboratory animals. III. Pathology of the experimental lesions in the rat. IV. Cultivation of the new mycobacterium. *LPB LX*: 93-122.

Maciag,A., Dainese,E., Rodriguez,G.M., Milano,A., Provvedi,R., Pasca,M.R. *et al.* (2007) Global analysis of the *Mycobacterium tuberculosis* Zur (FurB) regulon. *J Bacteriol* **189**: 730-740.

Maciag,A., Piazza,A., Riccardi,G., and Milano,A. (2009) Transcriptional analysis of ESAT-6 cluster 3 in *Mycobacterium smegmatis*. *BMC Microbiol* **9**: 48.

Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C., and Stover,C.K. (1996) Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol* **178**: 1274-1282.

McLaughlin,B., Chon,J.S., MacGurn,J.A., Carlsson,F., Cheng,T.L., Cox,J.S., and Brown,E.J. (2007) A mycobacterium ESX-1-secreted virulence factor with unique requirements for export. *PLoS Pathog* **3**: e105.

Ogura,T., Whiteheart,S.W., and Wilkinson,A.J. (2004) Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases. *J Struct Biol* **146**: 106-112.

Okkels,L.M., Brock,I., Follmann,F., Agger,E.M., Arend,S.M., Ottenhoff,T.H. *et al.* (2003) PPE protein (Rv3873) from DNA segment RD1 of *Mycobacterium tuberculosis*: strong recognition of both specific T-cell epitopes and epitopes conserved within the PPE family. *Infect Immun* **71**: 6116-6123.

Okkels,L.M., and Andersen,P. (2004) Protein-protein interactions of proteins from the ESAT-6 family of *Mycobacterium tuberculosis*. *J Bacteriol* **186**: 2487-2491.

Pajon,R., Yero,D., Lage,A., Llanes,A., and Borroto,C.J. (2006) Computational identification of beta-barrel outer-membrane proteins in *Mycobacterium tuberculosis* predicted proteomes as putative vaccine candidates. *Tuberculosis (Edinb)* **86**: 290-302.

Pallen,M.J. (2002) The ESAT-6/WXG100 superfamily – and a new Gram-positive secretion system? *Trends Microbiol* **10**: 209-212.

Parish,T. and Stoker,N.G. (2000) Use of a flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis tlyA plcABC* mutant by gene replacement. *Microbiol* **146**:1969-1975.

Philipp,W.J., Nair,S., Guglielmi,G., Lagranderie,M., Gicquel,B. and Cole,S.T. (1996) Physical mapping of *Mycobacterium bovis* BCG pasteur reveals differences in the genome map of *Mycobacterium tuberculosis* H37Rv and from *M. bovis*. *Microbiol* **142**: 3135-3145.

Poulet,S., and Cole,S.T. (1995) Characterization of the highly abundant polymorphic GC-rich-repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. *Arch Microbiol* **163**: 87-95.

Pym,A.S., Brodin,P., Majlessi,L., Brosch,R., Demangel,C., Williams,A. *et al.* (2003) Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nat Med* **9**: 533-539.

Quadri,L.E. (2008) Iron uptake in mycobacteria. In The mycobacterial cell envelope. Eds. Daffe,M., and Reyrat,J.M. (ASM Press, Washington DC) pp. 167-184.

Raghavan,S., Manzanillo,P., Chan,K., Dovey,C., and Cox,J.S. (2008) Secreted transcription factor controls *Mycobacterium tuberculosis* virulence. *Nature* **454**: 717-721.

Renshaw,P.S., Panagiotidou,P., Whelan,A., Gordon,S.V., Hewinson,R.G., Williamson,R.A., and Carr,M.D. (2002) Conclusive evidence that the major T-cell antigens of the *Mycobacterium tuberculosis* complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6*CFP-10 complex. Implications for pathogenesis and virulence. *J Biol Chem* **277**: 21598-21603.

Renshaw,P.S., Lightbody,K.L., Veverka,V., Muskett,F.W., Kelly,G., Frenkiel,T.A. *et al.* (2005) Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *EMBO J* **24**: 2491-2498.

Riley,R., Pellegrini,M. and Eisenberg,D. (2008) Identifying Cognate Binding Pairs among a Large Set of Paralogs: The Case of PE/PPE Proteins of *Mycobacterium tuberculosis. Plos Comp Biol* **4**: e1000174.

Rodriguez,G.M., Gold,B., Gomez,M., Dussurget,O., and Smith,I. (1999) Identification and characterization of two divergently transcribed iron regulated genes in *Mycobacterium tuberculosis*. *Tuber Lung Dis* **79**: 287-298.

Rodriguez,G.M., Voskuil,M.I., Gold,B., Schoolnik,G.K., and Smith,I. (2002) ideR, An essential gene in mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect Immun* **70**: 3371-3381.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) Molecular Cloning: A Laboratory Manual. 2nd Ed. (Cold Spring Harbour, New York).

Sampson,S.L., Lukey,P., Warren,R.M., van Helden,P.D., Richardson,M., and Everett,M.J. (2001) Expression, characterization and subcellular localization of the *Mycobacterium tuberculosis* PPE gene Rv1917c. *Tuberculosis (Edinb)* **81**: 305-317.

Sassetti,C.M., Boyd,D.H., and Rubin,E.J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**: 77-84.

Serafini,A., Boldrin,F., Palu,G., and Manganelli,R. (2009) Characterization of a *Mycobacterium tuberculosis* ESX-3 conditional mutant: essentiality and rescue by Iron and Zinc. *J Bacteriol* **191**: 6340-6344.

Shinnick,T.M., and Good,R.C. (1994) Mycobacterial taxonomy. *Eur J Clin Microbiol Infect Dis* **13**: 884-901.

Siegrist,M.S., Unnikrishnan,M., McConnell,M.J., Borowsky,M., Cheng,T-Y., Siddiqi,N., Fortune,S.M., Moody,D.B., and Rubin,E.J. (2009) Mycobacterial Esx-3 is required for mycobactin mediated iron acquisition. *PNAS* **106**: 18792-18797.

Simeone,R., Bottai,D. and Brosch,R. (2009) ESX/type VII secretion systems and their role in host-pathogen interaction. *Current Opinion Microbiol* **12:** 4-10

Singh,A., Mai,D., Kumar,A., and Steyn,A.J. (2006) Dissecting virulence pathways of *Mycobacterium tuberculosis* through protein-protein association. *Proc Natl Acad Sci U S A* **103**: 11346-11351.

Skjot,R.L., Oettinger,T., Rosenkrands,I., Ravn,P., Brock,I., Jacobsen,S., and Andersen,P. (2000) Comparative evaluation of low-molecular-mass proteins from *Mycobacterium tuberculosis* identifies members of the ESAT-6 family as immunodominant T-cell antigens. *Infect Immun* **68**: 214-220.

Snapper,S.B., Melton,R.E., Mustafa,S., Kieser,T., and Jacobs,W.R., Jr. (1990) Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* **4**: 1911-1919.

Sorensen,A.L., Nagai,S., Houen,G., Andersen,P., and Andersen,A.B. (1995) Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. *Infect Immun* **63**: 1710-1717.

Stanley,S.A., Raghavan,S., Hwang,W.W., and Cox,J.S. (2003) Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proc Natl Acad Sci U S A* **100**: 13001-13006.

Strong,M., Sawaya,M.R., Wang,S., Phillips,M., Cascio,D. and Eisenberg,D. (2006) Toward the structural genomics of complexes: Chrystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis. PNAS* **103:** 8060-8065.

Tekaia,F., Gordon,S.V., Garnier,T., Brosch,R., Barrell,B.G., and Cole,S.T. (1999) Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tuber Lung Dis* **79**: 329-342.

Teutschbein,J., Schumann,G., Mollmann,U., Grabley,S., Cole,S.T., and Munder,T. (2009) A protein linkage map of the ESAT-6 secretion system 1 (ESX-1) of *Mycobacterium tuberculosis*. *Microbiol Res* **164**: 253-259.

Tuberculist (2008) Tuberculist. URL http://genolist.pasteur.fr/TubercuList/.

Tundup,S., Akhter,Y., Thiagarajan,D., and Hasnain,S.E. (2006) Clusters of PE and PPE genes of *Mycobacterium tuberculosis* are organized in operons: evidence that PE Rv2431c is co-transcribed with PPE Rv2430c and their gene products interact with each other. *FEBS Lett* **580**: 1285-1293.

Tundup,S., Pathak,N., Ramanadham,M., Mukhopadhyay,S., Murthy,K.J., Ehtesham,N.Z., and Hasnain,S.E. (2008) The co-operonic PE25/PPE41 protein complex of *Mycobacterium tuberculosis* elicits increased humoral and cell mediated immune response. *PLoS One* **3**: e3586.

van der Wel,N., Hava,D., Houben,D., Fluitsma,D., van Zon,M., Pierson,J., Brenner,M. and Peters,P.J. (2007) *M. tuberculosis* and *M. leprae* translocate from the phagolysosome to the cytosol in myeloid cells. *Cell* **129**: 1287-1298.

van Pinxteren,L.A., Ravn,P., Agger,E.M., Pollock,J., and Andersen,P. (2000) Diagnosis of tuberculosis based on the two specific antigens ESAT-6 and CFP10. *Clin Diagn Lab Immunol* **7**: 155-160.

van Soolingen,D., van der Zanden,A.J.M., de Haas,P.E.W., Noordhoek,G.T., Kiers,A., Foudraine,N.A. *et al.* (1998) Diagnosis of *Mycobacterium microti* Infections among Humans by Using Novel Genetic Markers. *J Clin Microbiol* **36**: 1840-1845.

World Health Organization. (2004) BCG vaccine. WHO position paper, *Wkly Epidemiol Rec* **79**: 27–38.

World Health Organisation. (2007) WHO Report 2009: Global tuberculosis control - epidemiology, strategy, financing. WHO/HTM/TB.2007.376.

World Health Organisation (2008) Anti-tuberculosis drug resistance in the world, Fourth Global report. WHO/HTM/TB/2008.394.

World Health Organisation. (2009) WHO Report 2009: Global tuberculosis control - epidemiology, strategy, financing. WHO/HTM/TB/2009.411.

Wu,L.J. and Errington,J. (1994) *Bacillus subtilis* spoIIIE protein required for DNA segregation during asymmetric cell division. *Science* **264**: 572-575.

Xu,J., Laine,O., Masciocchi,M., Manoranjan,J., Smith,J., Du,S.J. *et al.* (2007) A unique Mycobacterium ESX-1 protein co-secretes with CFP-10/ESAT-6 and is necessary for inhibiting phagosome maturation. *Mol Microbiol* **66**: 787-800. L